Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade

Handbook for Practitioners

Christopher Paul, Jessica Yeats, Colin P. Clarke, Miriam Matthews, Lauren Skrabala

RAND CORPORATION

		Form Approved OMB No. 0704-0188
	Report Documentation Page	

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **2015**	2. REPORT TYPE	3. DATES COVERED **00-00-2015 to 00-00-2015**
4. TITLE AND SUBTITLE **Assessing and Evaluating DoD Efforts to Inform, Influence, and Persuade: Handbook for Practitioners**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **RAND Corporation,National Defense Research Institute,1776 Main Street, P.O. Box 2138,Santa Monica,CA,90407-2138**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**129**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

Cover image: A service member talks with a local official during outreach activities in the Philippines

(U.S. Army photo by 1st Lt. Ray Ragan).

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND's publications do not necessarily reflect the opinions of its research clients and sponsors.

Preface

This handbook distills observations, best practices, lessons, and recommendations tailored specifically to personnel charged with planning and assessing U.S. Department of Defense (DoD) efforts to inform, influence, and persuade. It was developed as part of the project "Laying the Foundation for the Assessment of Inform, Influence, and Persuade Efforts," which sought to identify and recommend selected best practices in assessment and evaluation drawn from existing practice in DoD, academic evaluation research, commercial marketing, public relations, public diplomacy, and public communication, including social marketing.

This handbook is intended to support practitioners charged with planning, executing, and assessing DoD efforts to inform, influence, and persuade, with its contents presented in a user-friendly, quick-reference format. A metaevaluation checklist designed for assessing actual influence efforts (though not for supporting or enabling efforts that do not have some form of influence as an outcome) is available for download with this handbook at http://www.rand.org/pubs/research_reports/RR809z2.html. An accompanying volume, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Desk Reference*, explores the points presented in this handbook in greater detail.[1] The contents of the desk reference target a wider range of stakeholders, serving as part advice to policymakers, part advice to assessment practitioners, and part reference guide on the subject.

This research was jointly sponsored by the Rapid Reaction Technology Office in the Office of the Under Secretary of Defense for Acquisition, Technology, and Logistics and the Information Operations Directorate in the Office of the Under Secretary of Defense for Policy. The research was conducted within the International Security and Defense Policy Center of the RAND National Defense Research Institute, a federally funded research and development center sponsored by the Office of the Secretary of Defense, the Joint Staff, the Unified Combatant Commands, the Navy, the Marine Corps, the defense agencies, and the defense Intelligence Community, under contract number W91WAW-12-C-0030.

For more information on the International Security and Defense Policy Center, see http://www.rand.org/nsrd/ndri/centers/isdp.html or contact the director (contact information is provided on the web page).

Contents, Visual Aids, and Abbreviations

Visual Aids

Figures

Tables

Boxes

Abbreviations

COA	course of action
DoD	U.S. Department of Defense
IE	information environment
IIP	inform, influence, and persuade
IO	information operations
IRC	information-related capability
FM	field manual
JOPP	joint operation planning process
JP	joint publication
MISO	military information support operations
MOE	measure of effectiveness
MOP	measure of performance
NATO	North Atlantic Treaty Organization
ROI	return on investment
SMART	specific, measurable, achievable, relevant, and time-bound
SME	subject-matter expert
TAA	target audience analysis

Summary

The U.S. Department of Defense (DoD) spends more than $250 million per year on information operations (IO) and information-related capabilities (IRCs) for influence efforts at the strategic and operational levels. How effective are those efforts? Are they well executed? How well do they support military objectives? Are they efficient (cost-effective)? Are some efforts better than others in terms of execution, effectiveness, or efficiency? Could some of them be improved? How? Unfortunately, generating assessments of efforts to inform, influence, and persuade (IIP) has proven to be challenging across the government and DoD. Challenges include difficulties associated with changes in behavior and attitudes, lengthy timelines to achieve impact, causal ambiguity, and struggles to present results in ways that are useful to stakeholders and decisionmakers.

This handbook addresses these challenges by reviewing and compiling existing advice and examples of strong practices in the defense sector, industry (including commercial marketing and public communication), and academia (evaluation research), drawn from a comprehensive literature review and more than 100 interviews with subject-matter experts across sectors. It then distills and synthesizes insights and advice for practitioners involved with planning and assessing DoD IIP efforts and programs.

> Begin with the end in mind.
>
> —Advice for social marketing campaigns (see Chapter Two)

An accompanying volume, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Desk Reference*, explores the points presented in this handbook in greater depth and detail.[1] The contents of the desk reference target a wider range of stakeholders, serving as part advice to policymakers, part advice to assessment practitioners, and part reference guide on the subject. This handbook further distills and synthesizes that content specifically for personnel charged with planning and assessing DoD IIP efforts.

---

[1] Christopher Paul, Jessica Yeats, Colin P. Clarke, and Miriam Matthews, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Desk Reference*, Santa Monica, Calif.: RAND Corporation, RR-809/1-OSD, 2015a.

How to Use This Handbook

This handbook was designed to be an easy-to-navigate, quick-reference guide to planning and conducting assessments of DoD IIP efforts, analyzing the data generated, and presenting the results to decisionmakers and stakeholders. As such, the layout is intended to provide the reader with a map to particular points of interest: The table of contents provides a complete breakdown of the chapters, topics, and accompanying visual aids, while Chapter One includes overview descriptions of each handbook chapter and a key throughout the handbook indicates the reader's place in the text. It also offers some background on current assessment practices in DoD, with connections to the joint operation planning process (JOPP), and the typical users and uses of DoD IIP assessment results. The discussion returns to these points repeatedly in the sections that focus on the assessment process and the presentation of assessment results. The need to balance thoroughness and conciseness means that not every possible topic is addressed here, and not every topic addressed here receives detailed treatment. The accompanying desk reference fills this gap for those who are interested in a more in-depth exploration or a wider range of examples. To help guide users to related topics here and in the desk reference, we offer suggestions for further reading throughout this handbook.

Good Assessment Practices Across Sectors

Across all the sectors in our study (industry, academia, and government), certain headline principles appeared again and again. We collected and distilled the most central (and most applicable to the defense IIP context). These are discussed in greater detail in Chapter Two.

Effective Assessment Requires Clear, Realistic, and Measurable Goals

How can you determine whether an effort has achieved its desired outcomes if the desired outcomes are not clear? How can you develop and design activities to accomplish desired goals if the desired goals have not yet been articulated? How can you evaluate a process if it is not clear what the process is supposed to accomplish? While the importance of setting clear goals may appear to be self-evident, too often, this obvious requirement is not met. Good assessment demands not just goals but clear, realistic, specific, and measurable goals.

Effective Assessment Starts in the Planning Phase

Assessment personnel need to be involved in IIP program planning to be able to point out when objectives are not specified in a way that can be measured and ensure that plans are made to make measurements and collect data. Likewise, planners need to be involved in assessment design to make sure that assessments will provide useful information and that they will have stakeholder buy-in. Building assessment into an IIP

effort from the very beginning also allows the impact of the effort to be tracked over time and enables failure to be detected early on, when adjustments and improvements can be made.

Effective Assessment Requires a Theory of Change or Logic of the Effort Connecting Activities to Objectives

Implicit in many examples of effective assessment and explicit in much of the work by scholars of evaluation is the importance of a *theory of change*. A theory of change, or logic of the effort, is the underlying logic for how planners think elements of an activity, line of effort, or operation will lead to desired results. Simply put, it is a statement of how you believe the things you are planning to do will lead to the objectives you seek. When a program does not produce all the expected outcomes and you want to determine why, a logic model (or other articulation of a theory of change) really shines.

> When a program does not produce all the expected outcomes and one wants to determine why, a logic model (or other articulation of a theory of change) really shines.
>
> —On the utility of logic models and theories of change (see Chapter Five)

Evaluating Change Requires a Baseline

While both the need for a baseline against which to evaluate change and the importance of taking a baseline measurement before change-causing activities begin seem self-evident, these principles are often not adhered to in practice. Without a baseline it is difficult to determine whether an IIP effort has had its desired impact—or any impact at all. You cannot evaluate change without a starting point.

Assessment over Time Requires Continuity and Consistency

Continuity and consistency are essential to the assessment of DoD IIP efforts. Behaviors and attitudes can change slowly over long periods, and data must be collected over the long term to provide an accurate picture of an effort's impact and to determine whether that impact was attributable to the effort itself or to some change in the context of the effort. If the data or the way they are collected changes during that time, it becomes harder to tell whether observed changes are due to changes in the behaviors or attitudes of interest or just to changes in how the behaviors are being measured. All military activities face a challenge in this area due to individual, unit, and command rotations, and IIP efforts are no exception.

Assessment Is Iterative

Assessment is an inherently iterative process, not something planned and executed once. Observing change over time requires repeated measurement over time. Further, it is unusual for an IIP effort to remain static for long, particularly in a complex environment. The context of an IIP effort can change, as can an effort's objectives or

the priorities of commanders and funders. Assessment must be able to adapt to these changes to help IIP efforts make course corrections.

> In a budget-constrained environment, evaluation is both more important and less affordable. You need a mechanism for quick, cheap, and "good enough" assessments.
>
> —Advice on designing high-quality assessments (see Chapter Seven)

Assessment Requires Resources

Organizations that routinely conduct successful and strong evaluations have a respect for research and evaluation ingrained in their organizational cultures, and they dedicate substantial resources to the conduct of evaluation. Unfortunately, assessment of DoD IIP efforts has been perennially underfunded. That said, some assessment (done well) is better than no assessment. Even if the scope is narrow and the assessment effort is underfunded and understaffed, any assessment that reduces the uncertainty under which future decisions are made adds value. And not all assessment needs to be at the same level of depth or quality. Where assessment resources are scarce, they need to be prioritized.

Challenges to Good Assessment and Successful IIP Efforts

Making Causal Connections

Because of the many actions and voices affecting the information environment, it is often difficult to tell whether a certain behavioral change was actually *caused* by defense IIP efforts. Where effectiveness is paramount, causation does not matter, and correlation is sufficient; if the target audience does what you want, you may not care exactly why. However, for accountability purposes, causation does matter. Being able to claim that a certain program or capability caused a certain effect or outcome increases the likelihood that the capability will continue to be valued (and funded).

While attributing causation in the information environment can be challenging, it is never impossible. If assessments need to demonstrate causal connections, thoughtful assessment design at the outset of the process can allow them to do so. See Chapter Seven, especially Box 7.1.

Building a Shared Understanding of DoD IIP Efforts

In our interviews, congressional staffers touched on a challenge that is inherent to IIP efforts relative to conventional kinetic military capabilities: a lack of shared understanding about, or intuition for, what IIP capabilities do and how they actually work (including a limited understanding of the psychology of influence).

Military personnel and congressional staffers have good intuition when it comes to the combined-arms contributions of different military platforms and formations.

They also have a shared understanding of the force-projection capabilities of a bomber wing, for example, or a destroyer, an artillery battery, or a battalion of infantry.

However, shared understanding does not extend to most IRCs. Intuition (whether correct or not) has a profound impact on assessment and expectations for assessment. Where shared understanding is strong, heuristics and mental shortcuts allow much to be taken for granted or assumed away; where there is a lack of shared understanding about capabilities, everything has to be spelled out, because the assumptions are not already agreed upon.

Where shared understanding is lacking, assessments must be more thoughtful. The dots must be connected, with documentation to policymakers and other stakeholders explicitly spelling out what might be assumed away in other contexts. Greater detail and granularity become necessary, as do deliberate efforts to build shared understanding. Despite the potential burden of the demand to provide congressional stakeholders with more information about IIP efforts and capabilities to support their decision-making and fulfill oversight requirements, there are significant potential benefits for future IIP efforts. Greater shared understanding can not only potentially improve advocacy for these efforts but also strengthen the efforts themselves by encouraging more-rigorous assessments. See the discussion in Chapter Three.

> The plural of anecdote is not data. Qualitative data should be generated by rigorous social science methods.
>
> —Advice on the role of qualitative approaches to assessment (see Chapter Eight)

Confronting Constraints, Barriers, Disruptors, and Unintended Consequences

If potential disruptors are considered as part of the planning process, they can also be included in the measurement and data collection plan. Collecting information in a way that takes into account potential points of failure can both facilitate adjustments to the effort and help ensure that assessment captures the effort's progress as accurately as possible. If the effort is found to be unsuccessful, it may be that there was not, in fact, a problem with the objectives or the underlying theory but that the effort has just been temporarily derailed by outside circumstances.

In a complex environment, IIP efforts face obstacles that can also challenge good assessment practices. For this reason, it is particularly important for DoD IIP assessment to incorporate the principles of good assessment articulated earlier and to ensure that an effort can adapt to changes in context. See the discussion in Chapter Five.

Learning from Failure

DoD requires IIP assessment for accountability purposes, of course, but it also depends on assessment to support a host of critical planning, funding, and process requirements. Consequently, it is vitally important to determine as early as possible whether

certain activities are failing or have failed. The unique challenge facing IIP planners is that they must do so without suggesting that IO overall is a failure.

Assessment can directly support learning from failure, midcourse correction, and planning improvements.[2] In military circles, there is a tendency to be overoptimistic about the likely success of an effort and be reluctant to abandon pursuits that are not achieving desired results. For this reason, we address failure—strategies to prevent it and strategies to learn from it—throughout this handbook.

After-action review is a familiar and widely used form of evaluation that is dedicated to learning from both success and failure. It has a major shortcoming, however: It is retrospective and timed in a way that makes it difficult for campaigns that are going to fail to do so quickly. The principles of good assessment articulated earlier can help prevent program failure, but they can also detect imminent failure early on, saving precious time and resources. When IIP efforts involve unvalidated assumptions or other uncertainties, structure the efforts and the assessments to *fail fast*, and then learn, iterate, and improve. See the discussion in Chapter Five.

Recommendations

This handbook contains insights that are particularly useful for those charged with planning and conducting assessment; the companion volume, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Desk Reference*, offers an abundance of information that is relevant to other stakeholders, including those who make decisions based on assessments and those responsible for setting priorities and allocating resources for assessment and evaluation.[3]

Our recommendations for assessment practitioners echo some of the most important practical insights described in the key takeaways at the end of each chapter and in the conclusions at the end of this handbook:

- *Demand specific, measurable, achievable, relevant, and time-bound (SMART) objectives.* Where program and activity managers cannot provide assessable objectives, assessment practitioners should infer or create their own.
- *Be explicit about theories of change/logic of efforts.* Theories of change ideally come from commanders or program designers, but, if the logic of an effort is not made explicit, assessment practitioners should elicit or develop one in support of assessment.

---

[2]   These three aims were emphasized, respectively, in an interview with Mary Elizabeth Germaine, March 2013; Marla C. Haims, Melinda Moore, Harold D. Green, and Cynthia Clapp-Wincek, *Developing a Prototype Handbook for Monitoring and Evaluating Department of Defense Humanitarian Assistance Projects*, Santa Monica, Calif.: RAND Corporation, TR-784-OSD, 2011, p. 2; and an interview with LTC Scott Nelson, October 10, 2013.

[3]   Paul et al., 2015a.

- *Insist that resources are provided for assessment.* Assessment is not free, and if its benefits are to be realized, it must be resourced. Presenting assessment results in ways that are tailored to specific stakeholders, highlighting successes in saving time and resources, and ensuring that data collection, measures, and results are as transparent as possible will help gain buy-in from stakeholders and DoD leadership.
- *Take care to match the design, rigor, and presentation of assessment results to the intended uses and users.* Assessment supports decisionmaking, and providing the best decision support possible should remain at the forefront of practitioners' minds. The ways in which assessment results will be used by decisionmakers must be a consideration throughout the assessment process. This may involve some amount of prediction, as decisionmakers may not always know what information they require, and it can be time-consuming and expensive to assemble the results required after data have been collected.

> It is important to do good science; it is also important to sell good science.
>
> —Advice on combining quantitative and qualitative data in presenting assessment results (see Chapter Eleven)

Practitioners depend to a great extent on leadership support and shared understanding with stakeholders and decisionmakers, just as leadership and stakeholders depend on practitioner understanding of their needs and resource constraints. As such, we reiterate some recommendations for the broader DoD IIP community, including stakeholders, proponents, and capability managers for IO, public affairs, military information support operations, and all other IRCs. The following recommendations, drawn from points in *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Desk Reference,* emphasize how advocacy and a few specific practices can improve the quality and use of assessment results across the community:

- DoD leadership needs to *provide greater advocacy, better doctrine and training, and improved access to expertise (in both influence and assessment)* for DoD IIP assessment efforts. Assessment is important for both accountability and improvement, and it needs to be treated as such.
- DoD doctrine needs to *establish common assessment standards.* There is a large range of possible approaches to assessment, with a similarly large range of possible assessment rigor and quality. The routine and standardized employment of something like the assessment metaevaluation checklist that accompanies this handbook online would help ensure that all assessments meet a target minimum threshold.
- DoD leadership and guidance need to recognize that *not every assessment must be conducted to the highest standard.* Sometimes, good enough really is good enough,

and significant assessment expenditures cannot be justified for some efforts, either because of the low overall cost of the effort or because of its relatively modest goals.

- DoD should *conduct more formative research*. Formative research can improve IIP efforts and programs and facilitate the assessment process. We offer the following specific recommendations:
  - Conduct target-audience analysis with greater frequency and intensity, and improve capabilities in this area.
  - Conduct more pilot testing, more small-scale experiments, and more early efforts to validate a specific theory of change in a new cultural context.
  - Try different things on small scales to learn from them (i.e., fail fast).
- DoD leaders need to *explicitly incorporate assessment into orders*. If assessment is in the operation order, the execute order, or even a fragmentary order, then it is clearly a requirement and will be more likely to occur, with requests for resources or assistance less likely to be resisted.
- DoD leaders should support the *development of a clearinghouse of validated (and rejected) IIP measures*. When it comes to assessment, the devil is in the details. Even when assessment principles are adhered to, some measures just do not work out, either because they prove hard to collect or because they end up being poor proxies for the construct of interest. Assessment practitioners should not have to develop measures in a vacuum. A clearinghouse of measures tried (with both success and failure) would be an extremely useful resource.[4]

---

[4] Paul et al., 2015a.

Acknowledgments

CHAPTER ONE
About This Handbook

This project's sponsors in the Office of the Secretary of Defense asked RAND to identify effective principles and best practices for the assessment of inform, influence, and persuade (IIP) efforts from across sectors and distill them for future application in the U.S. Department of Defense (DoD). As part of this effort, the RAND project team was asked to review existing DoD IIP assessment practices (and broader DoD assessment practices), identify IIP assessment practices in industry (commercial marketing, public relations, and public communication), and review guidance and practices from the academic evaluation research community.

To complete these tasks and provide DoD with a structured set of insights, principles, and practices applicable to the assessment and evaluation of IIP efforts, we conducted a comprehensive literature review and more than 100 interviews with subject-matter experts (SMEs) who held a range of roles in government, industry, and academia. The literature reviewed was copious and wide-ranging, encompassing hundreds of documents; we compiled the most informative and useful of those resources into an annotated bibliography and reading list, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: An Annotated Reading List*.[1] Many of our SME interviews were conducted on a for-attribution basis, so we are able to provide direct quotes and give credit where credit is due for good ideas.

We compiled the practices, principles, advice, guidance, and recommendations, distilling and synthesizing them for application to DoD in the form of a general reference, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Desk Reference*.[2] This handbook further distills and synthesizes that content, presenting it in a quick-reference format, and is intended specifically for personnel charged with planning and assessing DoD IIP efforts.

---

[1] Christopher Paul, Jessica Yeats, Colin P. Clarke, and Miriam Matthews, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: An Annotated Reading List*, Santa Monica, Calif.: RAND Corporation, RR-809/3-OSD, 2015b.

[2] Christopher Paul, Jessica Yeats, Colin P. Clarke, and Miriam Matthews, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Desk Reference*, Santa Monica, Calif.: RAND Corporation, RR-809/1-OSD, 2015a.

To keep the content streamlined and ensure its utility as a quick reference, this handbook cites just a small selection of applicable interviews and documents. Users of this handbook are encouraged to consult the desk reference for additional context, discussion, and examples. Consequently, we have included cross-references to other points in this handbook and in the accompanying desk reference where readers can find more-detailed discussions of various topics of interest.

The Language of Assessment

One factor that varies across government, defense, industry, and academia is *how* assessment is discussed. Different sectors use different terms of art to describe things that are similar, if not entirely overlapping. In government and defense, the term of choice is *assessment*, while academic evaluation researchers (unsurprisingly) talk about *evaluation*. In commercial marketing, the conversation is usually about *metrics* or just *measurement*. Others have written about *monitoring*, and many of the people we interviewed used more than one of these terms, sometimes as synonyms and sometimes to denote slightly different things. As one of these SMEs noted, "There are as many different definitions of *assessment* as there are people doing it."[3]

Here, we use *assessment* and *evaluation* interchangeably and synonymously, with our choice of the two terms driven by the source of the discussion: When the sources we are citing discussed *evaluation*, we use *evaluation*, and vice versa. When in doubt, or when the same topic was discussed by experts in multiple fields using different terminology, we lean toward *assessment* because it is the preferred term of art in the defense community. Where we use other terms (such as *measurement*, *measures of effectiveness*, or *formative evaluation*), we do so intentionally and specifically, and we make clear what we mean by those terms.

Outline of This Handbook

This handbook is structured in a way that roughly follows the assessment planning process, with background and recommendations for overall best assessment practices and the presentation of assessment results serving as bookends to topical discussions of planning assessments for decisionmaking, identifying objectives and selecting theories of change, developing measures, designing and implementing assessments, collecting data, and presenting and using assessment results.

---

[3]   Author interview on a not-for-attribution basis, December 5, 2012.

1 | **About This Handbook**
An introduction to the study describing this research and the structure for navigating the handbook.

2 | **Assessment Best Practices and Applying Them to DoD IIP Efforts**
Core assessment principles and DoD current practice identified and distilled.

3 | **Why Evaluate? An Overview of Assessment and Its Uses**
The overriding question driving this study—Why evaluate?—and the groundwork for the discussions to follow.

4 | **Determining What's Worth Measuring: Objectives**
Ideal properties for objectives to assess against, as well as best practices for the development and articulation of both objectives and logic models.

5 | **Determining What's Worth Measuring: Theories of Change and Logic Models**
Major theories of influence and persuasion that could inform the theory of change or logic for an IIP effort or program.

6 | **Developing Measures for DoD IIP Efforts**
Key concepts and best practices in developing the measures that can and should be used to evaluate the performance and effectiveness of IIP efforts.

7 | **Designing and Implementing Assessments**
Evaluation and assessment design, including criteria to help select the appropriate design.

8 | **Formative and Qualitative Research Methods for DoD IIP Efforts**
Data collection methods for formative evaluation and qualitative data collection methods more broadly.

9 | **Surveys and Sampling in DoD IIP Assessment: Best Practices and Challenges**
The use of surveys in IIP assessment, as well as survey sampling frames.

10 | **Measurement: Evaluating IIP Outputs, Outcomes, and Impacts**
Methods and data sources for assessing outputs, outcomes, and impacts (those appropriate or related to process and summative evaluation).

11 | **Presenting and Using Assessments**
Presenting assessments to maximize their utility and their ability to support decisionmaking.

12 | **Developing a Culture of Assessment**
How to organize for assessment.

13 | **Conclusions and Recommendations**
Connecting the dots with a focus on improvement.

Metaevaluation Checklist
A metaevaluation checklist for DoD IIP assessments accompanies this report on RAND's website at http://www.rand.org/pubs/research_reports/RR809z2.html.

Integrating Best Practices with Future DoD IIP Assessment Efforts: Operational Design and JOPP as Touchstones

Joint Publication (JP) 5-0, *Joint Operation Planning*, addresses both operational design and the joint operation planning process (JOPP).[4] While both are clearly aimed at a command staff during advance planning, they are sufficiently flexible to support a wide range of planning processes. Because JP 5-0 guidance is so broadly applicable and widely familiar to DoD personnel, we use operational design and JOPP throughout this handbook as touchstones to illustrate how and where the various assessment practices we recommend can be integrated into existing military processes. For those unfamiliar with operational design and JOPP, we briefly review both here.

Operational Design

As described in JP 5-0, operational art is about describing the military end state that must be achieved (ends), the sequence of actions that are likely to lead to those objectives (ways), and the resources required (means). This specification of *ends*, *ways*, and *means* sounds very much like the articulation of a theory of change (as described in Chapter Five).

Operational design is the part of operational art that combines an understanding of the current state of affairs, the military problem, and the desired end state to develop the operational approach. These are the four steps in operational design:

1. Understand the strategic direction.
2. Understand the operational environment.
3. Define the problem.
4. Use the results of steps 1–3 to develop a solution, i.e., the operational approach.

Joint Operation Planning Process

Operational design and JOPP are related in that operational design provides an iterative process that can be applied within the confines of JOPP. JOPP formally has seven steps: (1) planning initiation, (2) mission analysis, (3) course-of-action (COA) development, (4) COA analysis and war-gaming, (5) COA comparison, (6) COA approval, (7) plan or order development.

For practical purposes, mission analysis should be disaggregated so that it begins with a subprocess related to operational art—problem framing and visualization—and incorporates a full iteration of operational design. In our discussion of JOPP, we treat those two subprocesses as part of step 2, mission analysis. Those who would like further detail on either operational design or JOPP are referred to JP 5-0.

---

[4]   U.S. Joint Chiefs of Staff, *Joint Operation Planning*, Joint Publication 5-0, Washington, D.C., August 11, 2011a.

Assessment Best Practices and Applying Them to DoD IIP Efforts

Above all else, *assessment must support decisionmaking*, whether to inform campaign planning or execution, to help Congress enforce accountability for DoD activities, or to guide resource allocation decisions. Given what is at stake for DoD IIP programs, it is critical that practitioners adhere to the best available practices for planning and implementing assessments. Across all the sectors in our study (industry, academia, and government), certain headline principles appeared again and again. Here, we discuss each principle and what it looks like in practice in the context of DoD IIP efforts and assessments.

Assessment Best Practices

Effective Assessment Requires Clear, Realistic, and Measurable Goals

It appears to be self-evident that it is impossible to do assessment without having a clear goal in mind. Assessment and evaluation advice from every sector comes with an admonition to set clear goals. "Begin with the end in mind" is the advice given by Sarah Bruce and Mary Tiger for social marketing campaigns.[1] Too often this obvious requirement is not met. One DoD SME described defense IIP goals as "too often, lofty goals that are unattainable."[2] Assessment and evaluation require not just goals but clear, realistic, specific, and measurable goals. Goals must be realistic or assessment becomes unnecessary; unrealistic goals cannot be achieved, so there is no point in assessing. One defense SME we interviewed summed up the importance of clear, measurable objectives quite succinctly: "An effect that can't be measured isn't worth fighting for."[3]

The discussion of operational art and operational design in JP 5-0 highlights the importance of clear objectives while recognizing that complex or ill-defined problems or a disconnect between strategic and operational points of view can impede progress toward clear objectives. JP 5-0 notes, "Strategic guidance addressing complex problems can initially be vague, requiring the com-

---

[1] Sarah Bruce and Mary Tiger, *A Review of Research Relevant to Evaluating Social Marketing Mass Media Campaigns*, Durham, N.C.: Clean Water Education Partnership, undated, p. 3.

[2] Author interview on a not-for-attribution basis, July 30, 2013.

[3] Author interview on a not-for-attribution basis, December 5, 2012.

mander to interpret and filter it for the staff."[4] It goes on to note that subordinates should be aggressive in sharing their perspectives with higher echelons, working to resolve differences at the earliest opportunity. This is useful advice for assessors: If the objectives provided are too vague to assess against, try to define them more precisely and then push them back to higher levels for discussion and confirmation.

In JOPP, most of the elements of operational design should take place as part of step 2, mission analysis.[5] During mission analysis is when objectives should be articulated and refined, in concert with higher headquarters, if necessary. Clear objectives should be an input to mission analysis, but if they are not, mission analysis should provide an opportunity to seek refinement.

Effective Assessment Starts in Planning

Goal refinement and specification should be important parts of the planning process, and the need to articulate assessable goals and objectives is certainly part of what is meant when experts advise that assessment starts in planning. If poorly specified or ambiguous objectives survive the planning process, both assessment and mission accomplishment will be in jeopardy.[6]

There is more to it than that, however. In addition to specifying objectives in an assessable way during planning, assessments should be designed and planned alongside the planning of activities so that the data needed to support assessment can be collected as activities are being executed. Knowing what you want to measure and assess at the outset clarifies what success should look like at the end and allows you to collect sufficient information to observe that success (or its lack).[7]

Assessment personnel need to be involved in planning to be able to point out when an objective or subordinate objective is or is not specified in a way that can be measured and to identify decisions or decision points that could be informed by assessment. Assessors should involve planners in assessment design to ensure that assessments will provide useful information, that they will be designed to collect the desired data, and that they have stakeholder buy-in.[8]

LTC Scott Nelson, who served as the chief of influence assessment at U.S. Northern Command (USNORTHCOM), went so far as to suggest that "assessment should drive the planning process."[9] He argued that military planning and decisionmaking processes are designed in a way that supports assessment-driven planning: These

---

[4]   U.S. Joint Chiefs of Staff, 2011a, p. III-3.

[5]   See the section "Joint Operation Planning Process," in Chapter One, for the full list of steps.

[6]   Author interview on a not-for-attribution basis, January 23, 2013.

[7]   Author interview with Rebecca Andersen, April 24, 2013.

[8]   Author interview with Gerry Power, April 10, 2013.

[9]   Author interview with LTC Scott Nelson, October 10, 2013.

processes are supposed to work backward from measurable objectives in much the same way as good assessment design. In the words of Marine Air-Ground Task Staff Force Training Program materials, "Assessment precedes, accompanies and follows all operations."[10]

In the JOPP framework, assessment considerations should be present at the earliest stages. Formative assessment may inform operational design during mission analysis. Preliminary assessment plans should be included in COA development and should be war-gamed along with other COA elements during COA analysis and war-gaming.

Effective Assessment Requires a Theory of Change or Logic of the Effort Connecting Activities to Objectives

Implicit in many examples of effective assessment and explicit in much of the work by scholars of evaluation is the importance of a *theory of change*.[11] The theory of change or logic of the effort for an activity, line of effort, or operation is the underlying logic for how planners think that elements of the overall activity, line of effort, or operation will lead to desired results. Simply put, a theory of change is a statement of how you believe that the things you are planning to do are going to lead to the objectives you seek. A theory of change can include logic, assumptions, beliefs, or doctrinal principles. The main benefit of articulating the logic of the effort in the assessment context is that it allows assumptions of any kind to be turned into *hypotheses*. These hypotheses can then be explicitly tested as part of the assessment process, with any failed hypotheses replaced in subsequent efforts until a validated, logical chain connects activities with objectives and objectives are met. Here is an example of a theory of change:

> Training and arming local security guards makes them more able and willing to resist insurgents, which will increase security in the locale. Increased security, coupled with efforts to spread information about improvements in security, will lead to increased perceptions of security, which will, coupled with the encouragement to do so, promote participation in local government, which will lead to better governance. Improved perceptions of security and better governance will lead to increased stability.

As is often the case with IIP objectives, the IIP portion (increased perceptions of security and increased participation in local government) of this theory of change is just one line of effort in an array of efforts connected to the main goal. The IIP portion is dependent on the success of other lines of effort—specifically, real increases in security.

---

[10] U.S. Marine Corps, *Assessment: MAGTF Staff Training Program (MSTP)*, MSTP Pamphlet 6-9, Quantico, Va.: Marine Air-Ground Task Force Staff Training Program, October 25, 2007, p. 1.

[11] In presentations of early results, we noticed that some uniformed stakeholders were uncomfortable with the phrase *theory of change*, suggesting that *theory* sounds too theoretical, too abstract, and impractical. While used in the academic literature and throughout this handbook, where the phrase *theory of change* might create confusion, we include an alternative term of art, *logic of the effort*.

This theory of change shows a clear, logical connection between the activities (training and arming locals, spreading information about improving security) and the desired outcomes, both intermediate (improved security, improved perceptions of security) and long-term (increased stability). The theory of change makes some assumptions, but those assumptions are clearly stated, so they can be challenged if they prove to be incorrect. Further, those activities and assumptions suggest things to measure: the performance of the activities (training and arming, publicizing improved security) and the ultimate outcome (change in stability), to be sure, but also elements of all the intermediate logical nodes, such as the capability and willingness of local security forces, change in security, change in perception of security, change in participation in local government, and change in governance. Evaluation researchers assert that measures often "fall out" of a theory of change.[12]

Articulated at the outset, during planning, a theory of change/logic of the effort can help clarify goals, explicitly connect planned activities to those goals, and support the assessment process.[13] A good theory of change will also capture possible unintended consequences or provide indicators of failure, things to help you identify where links in the logical chain have been broken by faulty assumptions, inadequate execution, or factors outside your control (disruptors).[14]

Evaluating Change Requires a Baseline

To see change (delta), you need a starting point, a baseline with which to compare and from which to measure change. Further, it is best to measure the baseline before your interventions—your IIP activities—begin.[15] While the need for a baseline against which to evaluate change and the importance of taking a baseline measurement before change-causing activities begin again seem self-evident, these principles are often not adhered to in practice. One defense SME noted that baselines were often omitted because of insufficient time and resources.[16] Another observed that, sometimes, baseline data are collected but forces end up revising the baseline, either because the objectives changed (moving target) or because the next rotation of forces began the assessment process anew.[17]

---

[12] The quote is from the authors' interview with Christopher Nelson, February 18, 2013; for more on this general principle, see William J. McGuire, "McGuire's Classic Input-Output Framework for Constructing Persuasive Messages," in Ronald Rice and Charles Atkin, eds., *Public Communication Campaigns*, Thousand Oaks, Calif.: Sage Publications, 2012.

[13] Author interview with Maureen Taylor, April 4, 2013.

[14] Author interview with Steve Booth-Butterfield, January 7, 2013.

[15] Author interview with Charlotte Cole, May 29, 2013.

[16] Author interview on a not-for-attribution basis, January 23, 2013.

[17] Author interview on a not-for-attribution basis, September 8, 2013.

Without a baseline measurement of some kind to inform expectations, it would be impossible to say whether DoD efforts actually had any impact. It is sometimes possible to complete post hoc baselines against which to assess, but it is best to collect baseline data at the outset. Also note that while a baseline is essential to evaluating change, it is not always imperative that baseline data be quantitative. Sometimes, qualitative baseline data (such as data from focus groups) can provide a sufficient baseline.[18]

Assessment over Time Requires Continuity and Consistency

The previous discussion touched on "moving target" problems, where either the objectives change or the baseline is redone. These challenges point to a broader assessment principle—namely, the importance of continuity and consistency. A trend line is useful only if it reports the trend in a consistently measured way and if data are collected over a long enough period to reveal a trend. Assessment of progress toward an objective is useful only if that objective is still sought. Consistent, mediocre assessments are better than great, inconsistent assessments in many contexts.[19]

A lack of continuity and consistency is a problem in industry and in evaluation research,[20] but not at the same scale as in the defense sector. The major culprit in the defense context is rotation, including personnel rotation, unit rotation, and rotation at the senior command (and combatant command) levels. The frequent turnover of analysts can threaten continuity in assessment.[21] Further, whole assessment processes are often scrapped when new units rotate in and take over operations.[22] Especially in a military context, objectives—even long-term objectives—will change periodically.

Thoughtful nested or subordinate objectives can help mitigate against changing objectives at the highest level, provided existing subordinate objectives remain constant and still nest within new capstone objectives. Loss of continuity when rotating units abandon existing assessment frameworks might be avoidable if assessment practice improved in general, and if the leaders of the subsequent unit were more willing to accept existing "good enough" assessment rather than starting fresh every time.[23]

---

[18] Author interview with Kavita Abraham Dowsing, May 23, 2013.

[19] Author interview on a not-for-attribution basis, August 1, 2013.

[20] Peter H. Rossi, Mark W. Lipsey, and Howard E. Freeman, *Evaluation: A Systematic Approach*, Thousand Oaks, Calif.: Sage Publications, 2004.

[21] P. T. Eles, E. Vincent, B. Vasiliev, and K. M. Banko, *Opinion Polling in Support of the Canadian Mission in Kandahar: A Final Report for the Kandahar Province Opinion Polling Program, Including Program Overview, Lessons, and Recommendations*, Ottawa, Ont.: Defence R&D Canada, Centre for Operational Research and Analysis, DRDC CORA TR 2012-160U, September 2012.

[22] Author interview on a not-for-attribution basis, August 1, 2013.

[23] Author interview on a not-for-attribution basis, April 3, 2013.

Assessment Is Iterative

Assessment must be an iterative process, not something planned and executed once. First, efforts to track trends over time or to track incremental progress toward an objective require repeated, iterative measurement. Second, assessment needs to be planned and conducted iteratively, as things change over time; objectives can change, available data (or the ease of collecting those data) can change, and other factors can change, and assessment must change with them. Third, and related, IIP efforts involve numerous dynamic processes and thus require dynamic evaluation. Context changes, understanding of the context changes, theories of change change, and activities change based on revisions to theories of change; assessments need to adapt to reflect all of these changes. As IIP activities change, measures must be recalibrated and corrected, iteratively, along the way.[24] Fourth, as activities expand, assessment needs to change and expand with them. Just about any assessment effort will require some iteration and change.

Assessment Requires Resources

Organizations that routinely conduct successful evaluations have a respect for research and evaluation ingrained in their organizational cultures, and they dedicate substantial resources to evaluation.[25] The statement that assessment requires resources warrants a caveat, however. Especially for small-scale IIP efforts, assessment investment has to be reasonable relative to overall program costs. One cannot and should not spend more on assessment than on the activities being assessed!

With that in mind, our reviews and interviews suggested two further subordinate principles. First, some assessment (done well) is better than no assessment. Even if the scope is narrow and the assessment effort is underfunded and understaffed, any assessment that reduces the uncertainty under which future decisions are made adds value. Second, not all assessment needs to be at the same level of depth or quality. Where assessment resources are scarce, they need to be prioritized. For example, deemphasize efforts with very modest objectives or expenditures. Some efforts are not particularly extensive or ambitious, and progress toward those modest objectives could be assessed holistically, just based on the expert opinions of those conducting the activities. With certain military-to-military engagements, engaging at all is a step in the right direction. In other places (and for other audiences), the relationship is much more mature and IIP objectives have progressed beyond initial engagement and connection. The former scenarios require minimal assessment effort and expense, while the latter certainly merit more-substantial evaluation.

Another way to prioritize scarce assessment resources is to intentionally assess one effort to a high standard while allowing other, similar efforts to receive fewer assessment resources. If the logic of the effort is similar across efforts and the rigor-

---

[24] Author interview with David Michaelson, April 1, 2013.

[25] Author interview with James Deane, May 15, 2013.

Box 2.1
Bottom Line Up Front: The Most Informative Results for DoD IIP Efforts Come from the Intersection of Academic Evaluation and Public Communications

While usable and useful lessons came from all the sectors reviewed, the best insights came from the intersection of public communication (particularly social marketing) and academia. When we say *best*, we mean best in terms of applicability to defense IIP assessment, methodological rigor, and being novel to defense assessment. Public communication provided the best analogy for defense IIP. In the for-profit sector, many assessment efforts and measures connected to sales, earnings, return on investment (ROI), or something else that is explicitly monetized, which tends to break the analogy with defense. In public communication, however, behavior or attitudinal change is sought (as in defense IIP)—often from at-risk, hard-to-reach, or other challenging audiences (again, as in defense IIP). Where public communication has been conducted according to the best practices of evaluation, it has achieved a very compelling combination of effective, thoughtful assessment and methodological rigor. This combination is rare in existing defense IIP assessment practice, but we believe that the core principles and best practices from top-quality assessment efforts in public communication provide an excellent template for defense.

Achieving key U.S. national security objectives demands that the U.S. government and DoD effectively and credibly communicate with and influence a broad range of foreign audiences. To meet this objective, it is important to measure the performance and effectiveness of activities aimed at informing, influencing, and persuading. Thorough and accurate assessments of these efforts guide their refinement, ensure that finite resources are allocated efficiently, and inform accurate reporting of progress toward DoD's goals.

---

ously assessed effort validates that logic, then the performance of the other efforts can be reasonably inferred based on less-intensive monitoring. By contrast, if resources were spread evenly across similar efforts, unless those resources were robust, assessment could be insufficient for all.

Additional Lessons for DoD IIP Efforts

DoD requires IIP assessment for accountability purposes, of course, but it also depends on assessment to support a host of critical planning, funding, and process requirements. Many IIP efforts involve uncertainty. When trying to influence a population to do something new and different in a new context, there are many unknowns that might slow, diminish, or disrupt the effort. Under such circumstances, one way to figure out what works and what does not is to try something and observe the results. The guiding principle here should be to *fail fast*. If you try something and early and frequent assessment reveals that it is not working, you can adjust, correct, or try something else entirely.

Assessment can directly support learning from failure, midcourse correction, and planning improvements.[26] In military circles, there is a tendency to be overoptimistic about the likely success of an effort, and there is a reluctance to abandon pursuits that

---

[26] These three aims were emphasized, respectively, in an author interview with Mary Elizabeth Germaine, March 2013; Marla C. Haims, Melinda Moore, Harold D. Green, and Cynthia Clapp-Wincek, *Developing a Prototype Handbook for Monitoring and Evaluating Department of Defense Humanitarian Assistance Projects*, Santa

are not achieving desired results. For this reason, we address failure—strategies to prevent it and strategies to learn from it—throughout this report. More to the point: Building an organizational culture that values assessment requires getting over the fear of the results.

JP 5-0 describes operational design as an iterative process. Iteration should occur not just during initial planning but also during operations as assumptions and plans are forced to change in response to constraints, barriers, disruptors, and unintended consequences. Operational design also advocates continuous learning and adaptation, and well-structured assessment supports that process.

Further Reading

In this handbook:

Chapter Three provides examples of iteration with respect to meeting the needs of users of assessment results.

Chapter Five discusses in greater detail how to identify and articulate a theory of change or logic of the effort (and how to express a theory of change as a logic model).

Chapter Seven, in the section "Criteria for High-Quality Evaluation Design: Feasibility, Validity, and Utility," discusses trade-offs when designing assessments in a budget-constrained environment.

In the accompanying desk reference:

Chapter Three reviews best practices for DoD IIP assessment in greater detail and includes additional examples.

Key Takeaways

- Effective assessment requires clear, realistic, and measurable goals. "An effect that can't be measured isn't worth fighting for," nor is one that cannot be achieved.
- Assessment must start in planning for two reasons: to ensure that data collection and analysis are part of the plan (rather than something to be done, possibly inadequately, after the fact) and because the goals to be assessed must be established during the planning process.
- Assessment requires an explicit theory of change, a stated logic for how activities should lead to the results desired. Assessment along an effort's chain of logic enables process improvement, makes it possible to test assumptions, and can tell evaluators why and how (that is, where on the logic chain) an unsuccessful effort is failing.
- To evaluate change, a baseline of some kind is required. While it is sometimes possible to construct a post hoc baseline, it is best to have baseline data before the activities to be assessed have begun.

Monica, Calif.: RAND Corporation, TR-784-OSD, 2011, p. 2; and an author interview with LTC Scott Nelson, October 10, 2013.

- Assessment over time requires continuity and consistency in both objectives and assessment approaches. Consistent mediocre assessments are more useful than great, inconsistent assessments.
- The biggest threat to continuity and consistency in the defense context is rotation. Setbacks occur when new commanders change objectives and when new units change subordinate objectives and start new assessment processes.
- Assessment is iterative. Rarely does anything work exactly as intended, and contextual conditions change. Iterative assessment can show incremental progress toward objectives and help plans, processes, procedures, and understanding evolve.
- Assessment is not free; it requires resources. However, some assessment is better than no assessment, and not every activity merits assessment at the same level.

Why Evaluate?
An Overview of Assessment and Its Uses

This chapter lays a foundation for the discussion of assessment and evaluation that follows by describing the possible motives for assessment. We begin by identifying the core reasons for assessment, as well as some arguably illegitimate motives for evaluation. We then address the specific arguments for improved assessment of DoD IIP efforts, clarifying both the requirement for assessment and its utility and benefits.

Three Motivations for Evaluation and Assessment: Planning, Improvement, and Accountability

Assessment or evaluation is fundamentally a judgment of merit against criteria or standards.[1] But for what purpose? To what end do we make these judgments of merit? This report draws on examples from government and military campaigns, industry (both commercial marketing and public communication), and academia, collected through more than 100 interviews and a rigorous literature review to inform its findings. Across these sectors, all motivations or proposals for assessment or evaluation aligned comfortably with one (or more) of three broad goals: to improve planning, improve effectiveness and efficiency, and enforce accountability.

Three Types of Evaluation: Formative, Process, and Summative

The three broad motivations for assessment (improve planning, improve effectiveness and efficiency, and support accountability) roughly correspond to three primary types of evaluation. These concepts are drawn from the academic literature, so we use the term *evaluation* in this discussion; however, the implication is the same regardless of context. Shown in Figure 3.1, the three types or stages of evaluation are formative evaluation, process evaluation, and summative evaluation:

- *Formative evaluation* occurs primarily during the planning stage, prior to the execution of IIP activities, and includes efforts designed to develop and

---

[1] Rossi, Lipsey, and Freeman, 2004.

test messages, determine baseline values, analyze audience and network charac-
teristics, and specify the logic by which program activities are devised to generate
influence, including barriers to behavioral change.

- *Process evaluation* determines whether the program has been or is being imple-
mented as designed, assesses output measures (such as reach and exposure), and
provides feedback to program implementers to inform course adjustments.
- *Summative evaluation*, including "outcome" and "impact" evaluation, is the
postintervention analysis to determine whether the program achieved its desired
outcomes or impact.

These types of evaluation can be characterized as *stages*, because they can be
undertaken one after the other in an inherently linked way and can be conceptually
integrated as part of a full range of assessment activities over the duration of a program
or campaign. In this way, each stage informs those that follow.

For example, imagine planning and conducting an IIP effort to promote democ-
racy in a country by encouraging participation in national elections, not unlike efforts
that have occurred in Iraq and Afghanistan as part of Operation Iraqi Freedom and
Operation Enduring Freedom, respectively. The formative stage could include a range
of activities. One might begin by examining the records of election participation pro-
motion programs in other countries or previous efforts in the current country. The
formative stage is a good time to identify a baseline; in this case, voter turnout in pre-
vious elections would be a good baseline, supplemented by information about regional
variation or variation by different demographic characteristics, if possible. If a base-

Figure 3.1
Characteristics of the Three Phases of IIP Evaluation

Formative evaluation	Process evaluation	Summative evaluation

Activities
  Focus groups
  In-depth interviews
  Secondary analysis
  Participant observation
	Activities
	  Implementation monitoring
	  (e.g., viewer logs, broadcast
	  schedule)
	
	  Effects monitoring
	  (e.g., sales data, visitation
	  data, interviews)
		Activities
		  Analyze survey data
		  Key informant interviews

Objectives, understand:
  Barriers to action
  Appropriate language
  Constellation of factors
	Objectives, understand:
	  Frequency of broadcasts
	  Potential audience reach
	  Preliminary data on effects
		Objectives, understand:
		  Level of effect
		  Degree of efficiency

Design program	Launch program	Program ends

SOURCE: Based on a handout provided during author interview with Thomas Valente, June 18, 2013.
RAND *RR809/2-3.1*

line is not available (perhaps it is the first election under a new democratic scheme, or perhaps data were not recorded during previous elections), formative research could include preliminary surveys of intention to vote. Based on existing data or data collected as part of formative research, you could identify groups least likely to participate and try to identify ways to increase their participation. Formative research could include focus groups with representatives from populations of interest to identify barriers to participation in elections. Draft election-promotion materials could be presented and tested in other focus groups, with feedback contributing to revisions. Formative research could include limited pilot testing of materials with real audiences, provided there is some mechanism in place to see how well they are working (such as observations, a small survey, or quick interviews after exposure to the materials).

With as much planning and preparation as possible informed by the formative research, the delivery of the effort (what would be called the *intervention* in the academic literature) can begin. At this point, process evaluation can also begin.

An important part of process evaluation is making sure that the things that are supposed to happen are happening—and in the way envisioned. Are contractors delivering on their contracts? Are program personnel executing tasks, and are those tasks taking the amount of time and effort planned for them? Are audiences actually receiving materials as planned? Process evaluation is not just about recording these inputs, activities, and outputs; it is also about identifying problems in delivery, the reasons for those problems, and how they might be fixed. If, for example, a television commercial promoting election participation is being broadcast but no one reports seeing it, process evaluation turns back toward the methods of formative evaluation to find out why. Perhaps the commercial is airing on one channel in a time slot when the vast majority of the potential audience tunes in to a very popular program on a different channel. Note that while additional assessment activities begin when delivery begins, formative research need not stop. In this example, monitoring the early results of the election promotion program's delivery may provide new information that informs adjustments to the plan in progress.

For election-participation promotion, the core of summative evaluation takes place at the end: Was voter turnout increased by the desired amount or not? There is more to it than that, however. Even getting the answer to that simple question requires earlier thought and planning. If there is no baseline against which to compare voter turnout (either from a previous election or through some kind of projection), then change in turnout cannot be calculated. If objectives did not specify the desired increase in turnout, an absolute value of turnout or change in turnout could be calculated, but it would be difficult to know whether that is sufficient. Furthermore, those responsible for oversight of the effort might want to know how much of the change in turnout is attributable to the effort. This is a question about causation—often a particularly challenging one in the IIP context—and it would also be part of summative evaluation. If such a question is to be answered in the summative phase, it

has to be considered from the outset: Some form of quasi-experimental design would need to have been planned and executed, perhaps a design in which one or more areas were excluded from program delivery (either for a time or entirely), with differences

---

Box 3.1
Nesting: The Hierarchy of Evaluation

The nested relationship among the three stages of evaluation offers a slightly different conceptual scheme for thinking about evaluation. "The hierarchy of evaluation" as developed by the evaluation researchers Peter Rossi, Mark Lipsey, and Howard Freeman is presented below.[a] The hierarchy divides potential evaluations and assessments into five nested levels. They are nested in that each higher level is predicated on success at a lower level. For example, positive results for cost-effectiveness (the highest level) are possible only if supported by positive results at all lower levels.

Summative evaluation
Supports effectiveness/efficiency improvement and accountability

Process evaluation
Supports effectiveness/ efficiency improvement

Formative evaluation
Supports planning

Assessment of cost-effectiveness	Level 5
Assessment of outcome/impact	Level 4
Assessment of process and implementation	Level 3
Assessment of design and theory	Level 2
Assessment of need for effort	Level 1

SOURCE: Adapted from Christopher Paul, Harry J. Thie, Elaine Reardon, Deanna Weber Prine, and Laurence Smallman, *Implementing and Evaluating an Innovative Approach to Simulation Training Acquisitions*, Santa Monica, Calif.: RAND Corporation, MG-442-OSD, 2006, Figure 7.1.
RAND *RR809/2-Box3.1*

These five levels roughly correspond to the three motives and three stages of evaluation already described. Working from the bottom of the hierarchy, needs assessment and assessment of design and theory both support planning and are part of formative evaluation. Assessment of process and implementation directly corresponds to process evaluation and contributes to improving effectiveness and efficiency. Assessment of outcome/impact and assessment of cost-benefit effectiveness are part of summative evaluation and can be applied both to efforts to improve efficiency and effectiveness and to efforts to enforce accountability.

This framework is described as a hierarchy because the levels nest with each other; solutions to problems observed at higher levels of assessment often lie at levels below. If the desired outcomes (level 4) are achieved at the desired levels of cost-effectiveness (level 5), then lower levels of evaluation are irrelevant. But what about when they are not?

When desired high-level outcomes are not achieved, information from the lower levels of assessment needs to be available and examined. For example, if an effort is not realizing its target outcomes, is that because the process is not being executed as designed (level 3) or because the theory of change is incorrect (level 2)? Evaluators encounter problems when an assessment scheme does not include evaluations at a sufficiently low level to inform effective policy decisions and diagnose problems. When the lowest levels of evaluation have been "assumed away," skipping lower-level evaluation steps is acceptable only if those assumptions prove correct. By then, it could prove exceptionally difficult and costly to revisit those levels.

[a] Rossi, Lipsey, and Freeman, 2004.

in planned or actual voting behavior between areas exposed to the program and areas that were not (controlling for differences between the areas, perhaps statistically). This process would indicate the portion of the change in voter turnout due to the program.

Although the stages of evaluation seem sequential, being listed one after the other, they overlap and feed back onto each other, and all require some planning from the outset to execute properly.

Further Reading

In this handbook:

Chapter Seven connects assessment design with the types of evaluation described here.
Chapter Eight presents a number of formative and qualitative research methods that may be useful for IIP assessment.

In the accompanying desk reference:

Chapter Two explores each of these evaluation types in greater detail.
Chapter Seven explores formative, process, and summative evaluation design, and it connects these evaluation types to general IIP campaign elements and the seven-stage psychological operations process in the section "Types or Stages of Evaluation Elaborated: Formative, Process, and Summative Evaluation Designs."

Uses and Users of Assessment

Getting assessment results into a form that is useful to the people who need them to make decisions is one of the biggest challenges of assessment. If assessment is to support decisionmaking, it must be tailored in its design and presentation to its intended uses and users, and that must be done in a timely fashion. After all, methodologically rigorous assessments that fail to inform the decisionmaker before a decision is made are pretty much useless. Doing these things successfully requires a clear understanding of who will use the assessment results and how. Field commanders, for example, will have a different set of questions than congressional leaders.[2]

Evaluation researchers Peter Rossi, Mark Lipsey, and Howard Freeman have found that, unfortunately, some sponsors commission research with little intention of using the results.[3] Poorly motivated assessments include those done simply for the purpose of saying that assessment has taken place, those done to justify decisions already made, and those done to satisfy curiosity without any connection to decisions of any kind.[4] For example, if the commander asks for assessment to justify his or her chosen COA *after* it has been selected rather than before (during COA development or during COA analysis and war-gaming), then it is not really an assessment.

While assessment can have a range of uses and users and serve a number of different purposes, it should always support decisionmaking of some kind. This founda-

---

[2]   Author interview with Monroe Price, July 19, 2013.

[3]   Rossi, Lipsey, and Freeman, 2004.

[4]   Author interviews on a not-for-attribution basis, February 20 and October 30, 2013.

tional view is represented—if not always emphasized—in the best practices across the sectors we investigated. Here, we review a few of the primary users of assessment results and briefly discuss their needs and expectations.

Requirement 1: Congressional Interest and Accountability

Congressional scrutiny is the main driver of the evolving assessment requirement, and congressional interest can represent a very real threat to DoD IIP efforts. Some in Congress are highly skeptical of the efficacy of DoD's IIP efforts and would consider substantially curtailing such efforts and diminishing related capabilities.[5] Legislative decisions to be supported by assessments concern funding and authority: Which, if any, information operations (IO) programs should be funded? What legislative and policy constraints should be placed on the conduct of IO? What future oversight and reporting will be required?

Congressional staffers indicated that they would like to see assessments connect to strategy and to the outcomes of efforts. Mused one, "Could we get 'extent to which they accomplish [theater security cooperation plan] goals'?" These staffers also expressed a need for IO assessments that were more *standardized*. The desire for standardization clearly connects to oversight decisions. Congressional stakeholders wanted to understand why some programs receive more resources than others, and they wanted to see which programs are particularly effective (or cost-effective) to inform resource allocation decisions. Finally, staffers wanted assessments to justify IO activities as appropriate pursuits for DoD. An underlying current in many recent congressional inquiries can be captured by the question, "Shouldn't the State Department be doing that?"[6]

Good assessment, then, can meet multiple stakeholder needs by demonstrating that an IIP effort is effective and also by explicitly measuring its contribution to broader defense objectives. Congressional staffers indicated that it is much more compelling to measure the contribution of an effort to legitimate defense objectives than to simply argue that it contributes.[7]

Requirement 2: Improve Effectiveness and Efficiency

In addition to the importance of assessment for meeting congressional accountability demands, DoD relies on assessment to improve the effectiveness and efficiency of all its programs. The current era of fiscal austerity has put pressure on budgets across DoD, and budgets for IIP efforts are no exception. Opportunities to increase the effectiveness, and cost-effectiveness, of such efforts cannot be missed. Similarly, assessment can help monitor the performance of processes. Assessment supports learning from failure,[8]

---

[5]   Author interview on a not-for-attribution basis, May 7, 2013.

[6]   Author interview on a not-for-attribution basis, May 7, 2013.

[7]   Author interview on a not-for-attribution basis, May 29, 2013.

[8]   Interview with Mary Elizabeth Germaine, March 2013.

Box 3.2
Challenge: Lack of Shared Understanding

Congressional staffers have good intuition when it comes to the combined-arms contributions of different military platforms and formations. While they may not know the exact tonnage of bombs or shells required to destroy a bridge, they certainly understand that bombs and shells can be used to destroy bridges. However, shared understanding does not extend to most IRCs. Congressional stakeholders (and, to be fair, many military personnel) do not necessarily have a shared understanding of the value of a leaflet drop, a radio call-in program, or a military information support operations (MISO) detachment with a loudspeaker truck.

Intuition (whether correct or not) has a profound impact on assessment and expectations for assessment. Where shared understanding is strong, heuristics and mental shortcuts allow much to be taken for granted or assumed away; when there is a lack of shared understanding about capabilities, everything has to be spelled out. Consider the value of a capability—its ROI. As one of the military officers we interviewed remarked, "No one ever asks what the ROI was for a carrier strike group."[a] Many of the benefits of such naval forces are easy to comprehend but hard to quantify. There is, however, a shared understanding of the benefits (e.g., strike, deterrence, mobility, security, sometimes in a nebulous sense) and an appreciation for their complexity. There is also recognition of the time-conditional value of such capabilities: A carrier strike group has little ROI in port but a great deal of value during a contingency.

The story is slightly different when it comes to the ROI of IO investments and capabilities. Our interviews and literature review reinforced the conclusion that this is due to a general lack of shared understanding of the benefits of these efforts and the fact that many of these efforts are transitory (i.e., a contracted information campaign). For these reasons, there may be greater pressure to demonstrate the value of IO efforts. As one IO officer lamented, "We're held to different standards."[b] This appears to be true.

Where shared understanding is lacking, assessments must be more thoughtful. The dots must be connected, with documentation to policymakers and other stakeholders spelling out explicitly what might be assumed away in other contexts. Greater detail and granularity become necessary, as do deliberate efforts to build shared understanding. Despite the potential burden of the demand to provide congressional stakeholders with more information about IIP efforts and capabilities to support their decisionmaking and fulfill oversight requirements, there are significant potential benefits for future IIP efforts. Greater shared understanding can not only potentially improve advocacy for these efforts but also strengthen the efforts themselves by encouraging more-rigorous assessments.

[a] Author interview on a not-for-attribution basis, August 1, 2013.
[b] Author interview on a not-for-attribution basis, October 28, 2013.

midcourse correction,[9] and planning improvements.[10] DoD requires IIP assessment for accountability purposes, of course, but it also depends on assessment to support a host of critical planning, funding, and process requirements.

Requirement 3: Aggregate IIP Assessments with Campaign Assessments
The final noteworthy requirement for DoD IIP assessment concerns the aggregation of assessments of individual IIP activities with larger campaign goals. The challenge here is twofold. First, the assessment of individual activities and programs does not necessarily connect to the assessment of overall campaigns or operations. It is a familiar dilemma

---

[9] Haims et al., 2011, p. 2.

[10] Interview with LTC Scott Nelson, October 10, 2013.

in campaign planning and execution: *You can win the battles but still lose the war; the operation can be a success, but the patient can still die. The whole is sometimes greater than the sum of its parts.* This implies a requirement for assessment at multiple levels—at the level of the individual programs and activities, to be sure, but also at the level of contribution to overall campaigns. Second, assessments of IIP efforts need to be aggregated with other military lines of operation as parts of whole campaigns. This is necessary not only to assess the contribution of IIP efforts to broader campaigns but also to better integrate such efforts into routine military planning and into the overall military assessment process, a process from which IO have often been excluded, historically.[11]

Further Reading

In this handbook:

Chapter Seven discusses user needs in the context of assessment design (including instructions for building a uses/users matrix) in the section "Designing Useful Assessments."

Chapter Eleven describes how to match the presentation of assessment results to user needs.

In the accompanying desk reference:

Chapter Two explores the unique needs of various stakeholders in more detail in the section "Requirements for the Assessment of DoD Efforts to Inform, Influence, and Persuade."

Chapter Seven discusses the role of assessment as a decision-support tool in the section "Designing Useful Assessments and Determining the 'Users and Uses' Context," which also features a populated users/users matrix for a notional IIP program (see Table 7.5).

Key Takeaways

- Formative, process, and summative evaluations have nested and connected relationships in which unexpected results at higher levels can be explained by thoughtful assessment at lower levels. This is captured in the hierarchy of evaluation.
- Good assessment supports and informs decisionmaking.
- There is a range of different uses for and users of assessment. Assessments need to be tailored to the needs of end users in both their design and their presentation.
- Assessment of IIP efforts for accountability purposes is complicated by a lack of shared understanding or intuition. Everyone can intuit the value of kinetic military capabilities, but this is not necessarily true for IIP. A result is greater uncertainty about the basic value of IIP efforts and an increased need for granularity and specificity in IIP assessment.
- In addition to accountability, the DoD assessment requirement supports the greater effectiveness and efficiency of IIP efforts. Some good efforts can undoubtedly be better, and some weaker efforts could be made better through assessment.
- You can win the battles but still lose the war; the operation can be a success, but the patient can still die. DoD IIP assessment must address many needs simultaneously: those of the individual efforts, those of broader campaigns, and the contribution of the former to the latter.

---

[11]   Author interview on a not-for-attribution basis, August 1, 2013.

Determining What's Worth Measuring
Objectives

This chapter focuses on goals and objectives, the foundation for both operational and assessment success. The discussion highlights the properties that objectives should have and offers advice for setting (or refining) objectives so that they will have these desirable properties. We then address the expression of a theory of change that connects activities with the properly articulated objectives of the effort. Defining (or refining) objectives in an assessable way and articulating a theory of change (or logic of the effort) are foundational for assessment success.

Setting objectives for an IIP effort or activity is a nontrivial matter. While it is easy to identify high-level goals that at least point in the right direction (e.g., "win," "stabilize the province," "promote democracy"), getting from ambiguous aspirations or end states to useful objectives is challenging. Yet clear objectives are necessary for not only the design and execution of effective IIP efforts but also their assessment. This section describes some of the challenges and tensions inherent in setting IIP objectives and offers some advice regarding considering and setting objectives.

Characteristics of SMART or High-Quality Objectives

The received wisdom on assessment holds that objectives should be "SMART"—that is, specific, measurable, achievable, relevant, and time-bound.[1] Table 4.1 summarizes each of these criteria; each is then explored in greater detail, along with a selection of additional virtues to which objectives should aspire.

Specific

How can you talk about progress toward or accomplishment of a goal if you have not specified what the goal really is? This is particularly important for IIP efforts and their assessment because objectives in this area need to, according to one SME, "be very literal." It can be a source of difficulty when objectives are "abstract or wishy-washy."[2]

---

[1] Author interview with Thomas Valente, June 18, 2013; Jessica M. Yeats and Walter L. Perry, *Review of the Regional Center Enterprise Measures of Effectiveness Plan*, unpublished RAND research, 2011, p. 9; author interview with Anthony Pratkanis, March 26, 2013.

[2] Interview with Emmanuel De Dinechin, May 16, 2013.

Table 4.1
Characteristics of SMART Objectives

An Objective Is . . .	If...
Specific	It is well defined and unambiguous and describes exactly what is expected
Measurable	One can measure the degree to which the objective is being met
Achievable	It is realistic and attainable
Relevant	The achievement of the objective contributes to progress toward high-level strategic and policy goals
Time-bound	It has deadlines or is grounded within a deadline

SOURCE: Yeats and Perry, 2011, p. 9.

IIP objectives need to specify what behavior or behavior change is desired and from what audience or group.[3] Army Field Manual (FM) 3-13, *Inform and Influence Activities*, presents a scheme for generating objective statements that, if followed, would certainly help a user meet the "specific" requirement. According to FM 3-13, an inform and influence objective statement should have four elements, each of which should be clearly articulated: the desired effect or outcome, the specific target, the desired target behavior, and the rationale for getting the target to perform that behavior (connecting the behavior to the outcome).[4] Figure 4.1 illustrates this construct.

It is important that objectives specify *what* is to be accomplished, not *how* it is to be accomplished. As noted in JP 5-0, "An objective does not infer ways and/or means—it is not written as a task."[5] Consider some of the objectives that correspond to the DoD IIP examples used in this report so far. The objective to promote voter

Figure 4.1
Sample Inform and Influence Activities Objective Statement
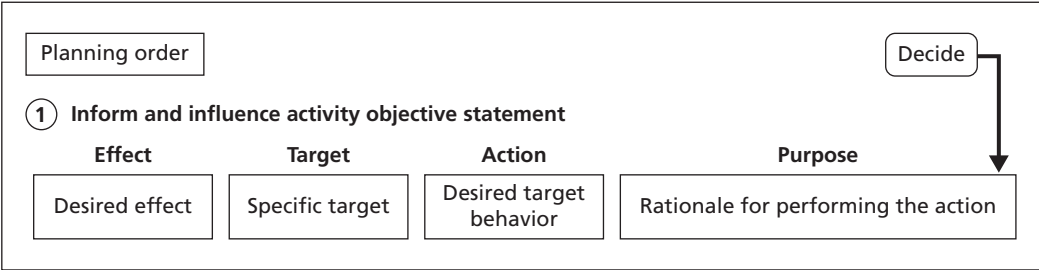
SOURCE: Headquarters, U.S. Department of the Army, 2013a, Figure 7-1.
RAND *RR809/2-4.1*

---

[3]   Interview with Anthony Pratkanis, March 26, 2013.

[4]   Headquarters, U.S. Department of the Army, *Inform and Influence Activities*, Field Manual 3-13, Washington, D.C., January 2013a, p. 7-2.

[5]   U.S. Joint Chiefs of Staff, 2011a.

turnout is fairly clear, but it could be more specific. The desired action is clear: Get the target audience to vote. The previous discussion made the purpose clear: Support democratization and governance processes. What is not clearly specified is the target audience, which could be all eligible partner-nation citizens or perhaps one or more traditionally underrepresented groups. The extent of the desired effects could also be better specified: Among the target audiences, what is the desired level of increased voter turnout? Five percent? Ten percent? Specificity to that level forces more-careful planning and encourages proactive refinement if interim measures show that the effort has not made as much progress as desired.

Measurable

A measurable objective is one that can be observed, either directly or indirectly. High-quality objectives will allow observation of the degree to which the objective is being met (percentage of population adopting desired behavior or frequency with which targeted audience engages in desired behavior) rather than all or nothing (extremist rhetoric eliminated from radio broadcasts).

Some objectives, even those that are not behavioral and cannot be directly observed, can still be meaningfully measured. Customer satisfaction is one example, as are various desired sentiments or attitudes. While perception of security cannot be directly observed, it can be self-reported in an interview, survey, or focus group, and it is likely to be highly correlated with proxy behaviors that *can* be directly observed. Pedestrian and vehicular traffic in an area, the number of people in the market on market day, and the percentage of school-age children who actually attend school are all observable and measurable things that could be proxy indicators for perceptions of security.

One way to move toward measurable objectives is to ask as part of the objective-setting process, "How will we know if we are meeting the objective?" If that question produces a clear idea about something to observe, or a clear indicator or measure to capture, then the objective is probably already measurable. If, on the other hand, that question prompts no clear answer, the objective should probably be refined.

Some objectives are just too complex or high level to be meaningfully observed directly, such as democratization or legitimacy. These are still worthwhile strategic goals, but they should be supported by measurable subordinate objectives (see the discussion of nesting in Box 3.1 in Chapter Three). Measure development is discussed in greater detail in Chapter Six.

Achievable

An objective must be something that one can reasonably expect to achieve. No IIP program is going to solve world hunger.[6] IO SMEs informed us that DoD IIP efforts are certainly not immune to this kind of objective inflation. Nor is public diplomacy.

---

6   Author interview on a not-for-attribution basis, July 30, 2013.

As the public diplomacy expert Phil Seib reminded us, "Success doesn't mean loving America." It is much more beneficial to set reasonable standards and benchmarks on objectives that are more realistic and useful.[7]

Achievable objectives are a balance between reasonable goals and reasonable expectations. Changing behaviors can require significant investments of time and resources, and it does not always work.[8] Those planning and executing IIP efforts must be patient and not expect to see immediate or extreme results. This is another area in which breaking objectives into smaller incremental chunks can be helpful, as the level of effort that turns out to be required to achieve the earliest and simplest of nested and progressive objectives can provide some indication of how difficult it will be to achieve subsequent objectives—if, in fact, the full scope of objectives is achievable in a reasonable time frame.

Goals can be unachievable in two ways: The goal could be impractical or the timeline for achieving it could be impossible. Getting 100-percent voter turnout or reducing the incidence of violence in a troubled province to zero is just not possible. Increasing voter turnout from 50 to 60 percent or reducing violent incidents from 50 per month to fewer than 15 per month might be possible but could not be accomplished in a single week. The SMART characteristics are mutually reinforcing; if objectives are specific, it is much easier to ascertain whether they are achievable or not.

Relevant

Nesting objectives such that they are clearly connected also helps ensure that objectives are relevant to overall end states or campaign goals. If one is not careful, it is entirely possible to specify objectives that are observable and measurable but not actually connected to the mission or desired end state. Irrelevant (but achievable) objectives are harder to avoid if the implied or explicit theory of change does not adequately connect intermediate or tactical objectives with campaign or long-term objectives. This is what happens in situations analogous to winning all the battles but losing the war. As JP 5-0 states, "An objective should link directly or indirectly to higher level objectives or to the end state."[9]

Irrelevant objectives are usually "missing a link" in their theory of change/logic of the effort. A defense SME shared an anecdote about a "tip line to nowhere."[10] In the country of interest, an IIP effort sought to persuade local citizens to report suspicious activity to a tip line. IIP activities were conducted, and a line was established. A few months after the effort began, the tip line began receiving a significant number of calls, and the effort was considered successful. However, while the effort met the stated

---

[7]   Author interview with Phil Seib, February 13, 2013.

[8]   Author interview with Larry Bye, June 19, 2013.

[9]   U.S. Joint Chiefs of Staff, 2011a.

[10]   Author interview on a not-for-attribution basis, March 13, 2014.

objective of changing local behavior to report suspicious activity to a tip line, it was not successful in any real sense. Why? Because the line was not "connected" to anything. That is, there was no procedure in place to validate the tips through other sources and then pass them to local authorities (or anyone else) to investigate or act on them. Tips were simply recorded in a logbook that then *just sat there*. The objective of collecting tips, was, by itself, not relevant to the campaign; only when and if collecting tips was connected to superordinate and longer-term objectives related to the reduction of criminal or insurgent behavior and the capture of perpetrators would it become relevant.

Time-Bound

Finally, an objective should include a time horizon for its completion. Objectives that are not time-bound invite efforts in perpetuity that are making little or no real progress. Even if the desired end state is a generational change in international relationships, the intermediate objectives should have some kind of indicated time scope. Time boundaries need not be more precise than the science will allow, and they can be phrased as opportunities to assess progress and revisit plans rather than times after which progress will be considered to be lagging. The timing of objectives can be tied to other natural temporal boundaries. How much progress on this chain of objectives do you think you will have made by the elections next year? How much progress on this objective will you make during your duty rotation? Timing should be specified, and so should the preliminaries of what should happen (be it taking a benchmark measure, some kind of scrutiny, revisiting the theory of change, launching the next phase of the effort, or considering canceling the activity) when a time boundary is reached.

Behavioral Versus Attitudinal Objectives

There is debate within the defense IIP community about whether objectives should be exclusively behavioral or whether attitudinal objectives are also permissible. The argument goes something like this: If influence is to contribute to military objectives, it will be because it gets people to do (or not do) certain things (engage in behaviors) that support broader military objectives. There is general agreement that changes in attitude *might* lead to the adoption of the desired behaviors; if you know what those desired behaviors are, you should specify them as part of the objective. For example, if the objective is reduced support for the insurgents, desired behavior changes might include decreased provision of havens to the insurgents, decreased provision of money or supplies to the insurgents, or decreased turnout at pro-insurgent demonstrations or protests. While many of these behaviors might correlate with or even stem from attitudes that are less supportive of the insurgency, the objective is really about the behaviors, even if changing attitudes is part of the planned effort.

However, where attitudes do *not* predict behavior well, the debate matters, and specifying behavioral objectives should be strongly preferred. Fortunately, articulating

a clear theory of change/logic of the effort that connects planned activities with desired end states (as we advocate) allows the specification of both attitudinal and behavioral intermediate objectives and allows them to be tested as hypotheses in context as part of assessment. If a theory of change specifies a path promoting, first, attitudinal change, then behavioral change, and then achievement of the desired end state, the validity of this path can be tested.

While we do not resolve this debate here, if the ultimate goal or end state requires that something demonstrable has changed (be it an adversary's capitulation, the election of a government friendly to the United States, or something else), it is probably best to specify the behaviors that will lead to those end states rather than stopping at attitudes favorable to those end states. And if (as we advocate) planners have specified a string of nested and progressive intermediate objectives, there is no harm (and there may be a benefit) in having these nesting objectives include a mix of attitudinal and behavioral elements. Again, behavioral objectives are strongly preferred over attitudinal objectives. Attitudinal changes may be included as subordinate or supporting objectives and as part of a longer chain of logic, but ultimate objectives should include some kind of consequential behavioral change.

Box 4.1
Setting Target Thresholds: How Much Is Enough?

A combination of the specific, achievable, and time-bound aspects of SMART informs the step of setting target thresholds for objectives. How much is enough? What proportion of a target audience needs to adopt a desired behavior for the effort to be considered a success? What level of progress do you need to make toward an intermediate objective before you launch activities that aim to build on that progress and before you move the effort toward accomplishing a later subordinate objective? At what threshold have your efforts accomplished all they can toward this objective, indicating that it is time to transition to different efforts and objectives or to take the program elsewhere?

Once again, your desired end state and ultimate goal should help drive thresholds. In an election, 51 percent voting for your preferred candidate is an unambiguous success.[a] However, for an effort promoting voter turnout, what amount of improvement is desired? Almost no IIP effort should expect 100-percent change or accomplishment, whatever the objective. Even where an objective is relative, seeking an increase or decrease in a behavior (such as "decrease insider attacks in province X"), it should be accompanied by a target threshold—expressed in percentage or absolute terms.

Another way to think about the target threshold is in a decisionmaking context. Remember that assessment should support decisionmaking. How much of something do you need to see to reach a decision point, or for you feel compelled to choose a different course of action?[b]

Clear target thresholds can help mitigate against open-ended commitments (where "improvement" continues to be sought long after enough of whatever was improving has been gained), and they can help turn "good enough" into "better" the next time by identifying weaknesses in theory or practice. An effort should have termination criteria—clear guidelines for what constitutes sufficient accomplishment to move on to the next stage of the effort or to consider the effort complete.[c]

[a] Author interview with Mark Helmke, May 6, 2013.

[b] Douglas W. Hubbard, *How to Measure Anything: Finding the Value of "Intangibles" in Business*, Hoboken, N.J.: John Wiley and Sons, 2010.

[c] U.S. Joint Chiefs of Staff, *Commander's Handbook for Assessment Planning and Execution*, version 1.0, Suffolk, Va.: Joint and Coalition Warfighting, September 9, 2011b.

Intermediate Versus Long-Term Objectives

Related to the time-bound aspect of SMART objectives is the potential tension between intermediate and long-term objectives. Many IIP end states are long-term and do not lend themselves to intermediate measures of progress.[11]

The solution, of course, is to have both intermediate and long-term objectives. Specify the long-term objective as precisely as possible and keep it available as a constant reference. Then, identify the incremental steps that you believe will lead you to that end state: "Define what conditions will change at each phase and how to detect the new behavior or function."[12] These intermediate objectives provide actionable and assessable objectives in the short- and medium-terms. Further, beliefs about the steps necessary to reach a desired end state can be tested as hypotheses. Does the second intermediate objective actually lead to the third intermediate objective? If not, revise it (sooner rather than later) so that a solid logical connection can still be made between intermediate objectives and the ultimate long-term objective.

For example, the ultimate objective for the tip line mentioned earlier could have been to take action against insurgents based on synthesis of citizen tips and corroborating intelligence, with a secondary objective to increase citizen participation in legitimate government processes, such as the reporting of criminal or insurgent behavior. Intermediate objectives, then, would include not only establishing and advertising the tip line but also transmitting tips received to relevant parties (such as law enforcement), the timely validation of tip intelligence, and timely action based on the tips.

How to Identify Objectives

Much of the discussion so far has focused on the characteristics of well-formed IIP objectives. Often, just identifying the desired characteristics will push a planner toward better-specified objectives. However, it is sometimes the case that the overall goal is clear but how to describe the objectives effectively is not. In our research, we encountered a number of processes for identifying and refining objectives.

One piece of advice is to work with stakeholders to better refine goals and objectives. If initial guidance from higher levels is not sufficiently specific, return with clarifying questions: Who? What? How much? By when?[13] Even absent broad stakeholder engagement, these are good questions. If objectives are insufficiently articulated in guidance from the higher level, those at the planning and execution level can try to refine objectives until they are SMART. These refined objectives can then be pushed back up to the higher level for approval.

---

[11] Author interview on a not-for-attribution basis, August 1, 2013.

[12] The Initiatives Group, *Information Environment Assessment Handbook*, version 2.0, Washington, D.C.: Office of the Under Secretary of Defense for Intelligence, 2013, p. 21.

[13] Ketchum Global Research and Analytics, *The Principles of PR Measurement*, undated, p. 6.

The third chapter of JP 5-0, "Operational Art and Operational Design," urges commanders to collaborate with their higher headquarters to resolve differences in interpretation regarding objectives in order to achieve clarity. This should be done as part of the "understand the strategic direction" element of operational design, and it should take place in JOPP during the planning initiation or mission analysis step (or perhaps between them).

Further Reading

In this handbook:

Chapter Three addresses nesting in Box 3.1, "Nesting: The Hierarchy of Evaluation."

Chapter Five explores the role of objectives in theories of change and the development of logic models, as well as the process of working backward from SMART objectives (in the section "Find and Fill Gaps in the Logic Model").

In the accompanying desk reference:

Chapter Three builds on the nesting concept by connecting nested objectives to the IIP context, including how an overall objective can be broken into several subordinate, intermediate, or incremental steps.

Chapter Five, in the section "How to Identify Objectives," explains how to move on to objectives by first identifying values. The section "How IIP Objectives Differ from Kinetic Objectives" offers an overview of what makes IIP objectives unique.

Key Takeaways

- The quality of an effort's goals directly relates to the quality of its associated assessment measures. Clearly articulated and specific goals are much easier to connect to clear and useful measures.
- Good IIP objectives should specify the observable behaviors sought, and from whom they are sought (the target audience).
- While there is some debate, behavioral objectives are strongly preferred over attitudinal objectives. Attitudinal changes may be included as subordinate or supporting objectives and as part of a longer chain of logic, but ultimate objectives should be some kind of consequential behavioral change.
- Good objectives are SMART: specific, measurable, achievable, relevant, and time-bound.
- Good objectives need to at least imply what failure would look like. How will you know if you have not succeeded?
- Breaking objectives into smaller "bite-sized" incremental subordinate objectives can make it easier to articulate a logic model or theory of change and make it possible to demonstrate incremental progress.

Determining What's Worth Measuring
Theories of Change and Logic Models

One of the recurring themes of this report is the importance of (and the benefits from) specifying a theory of change or logic of the effort for an IIP effort. A logic model is one way to collect and express the elements of a theory of change: "The logic model is supposed to make the program's theory of change explicit. A theory of change describes how the activities, resources, and contextual factors work together to achieve the intended outcome."[1]

Logic Model Basics

Logic models traditionally include program or effort inputs, outputs, and outcomes. Some styles of logic model development also report activities and impacts. Figure 5.1 presents these elements in sequence.

Inputs, Activities, Outputs, Outcomes, and Impacts

The *inputs* to a program or effort are the resources required to conduct the program. These will of course include personnel and funding, but are usually more specific than this, perhaps indicating specific expertise required or the number of personnel (or person-hours of effort) available. An effort's *activities* are the verbs associated with the use of the resources, and they are the undertakings of the program; these might include the various planning, design, and dissemination activities associated with messages or products, and could also include any of the actions necessary to transform the inputs into outputs. In fact, some logic model templates omit activities, as activities just connect inputs to outputs and can often be inferred by imagining what has to be done with the inputs to generate the outputs. We include activities here because of the focus on informing, influencing, and persuading, and the fact that assumptions are not always shared, and there is certainly no harm in being explicit about what activities will transform the inputs into outputs.

The *outputs* are produced by conducting the activities with the inputs. Outputs include traditional measures of performance (MOPs) and indicators that

---

[1] Donna M. Mertens and Amy T. Wilson, *Program Evaluation Theory and Practice: A Comprehensive Guide*, New York: Guilford Press, 2012, p. 244.

Figure 5.1
Logic Model Template
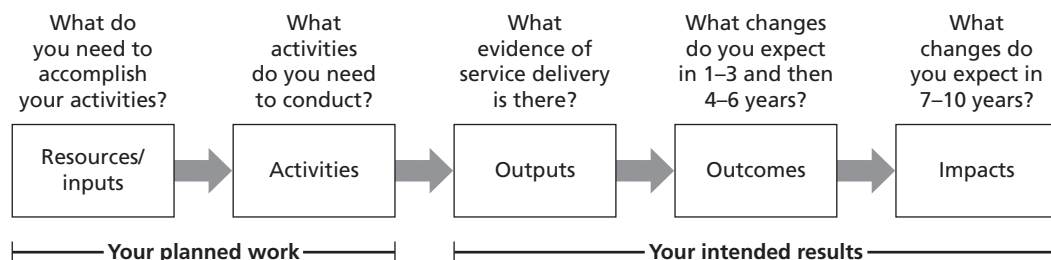
SOURCE: Mertens and Wilson, 2012, p. 245, Figure 7.1. Used with permission.
RAND *RR809/2-5.1*

the activities have been executed as planned. These might include execution and dissemination indicators, measures of reach, measures of receipt/reception, indicators of participation, and so on. *Outcomes* (or effects) are "the state of the target population . . . that a program is expected to have changed."[2] This is the result of the process: The inputs resource the activities, and the activities produce the outputs. The outputs lead to the outcomes. This is a critical juncture from a theory of change perspective, as the mechanism by which the outputs (messages disseminated, messages received) connect to the outcomes (behaviors changed) is critical and is a potentially vulnerable assumption in influence and persuasion. Outcomes are characteristics or behaviors of the audience or population, not of the program or effort. The outputs are related to the program or effort, and they describe the products, services, or messages provided by the program. Outcomes refer to the results (or lack of results) of the outputs produced, not just their delivery or receipt.[3]

The *impact* of a program or effort is the expected cumulative, long-term, or enduring contribution, likely to a larger campaign or superordinate goal. There is no clear dividing line between immediate and short-term outcomes, medium-term outcomes, and long-term impacts. In fact, there is no agreed-upon difference between outcome and impact. To some, this difference is one of individual change versus system change;[4] to others, it means a difference in design in that *outcomes* are not proven to be causally linked to the activities and outputs, but *impacts* are those outcomes that can be attributed to the intervention due to evidence from (typically) experimental studies.[5] To others, it is just a time horizon or level of analysis, with impacts being long-term

---

[2]   Rossi, Lipsey, and Freeman, 2004, p. 204.

[3]   Rossi, Lipsey, and Freeman, 2004.

[4]   Amelia Arsenault, Sheldon Himelfarb, and Susan Abbott, *Evaluating Media Interventions in Conflict Countries*, Washington, D.C.: United States Institute of Peace, 2011, p. 16.

[5]   Author interview with Julia Coffman, May 7, 2013.

and expanded outcomes.[6] Under this last scheme, if the outcome is the changing of a specific set of behaviors or attitudes, the impact is the durability of that change and the broader consequences of that change. For example, if the outcome of a defense IIP effort is increased participation in an election in a partner nation, the hoped-for impact might be a combination of increased participation in future elections and increased support for democracy and democratic values.

JP 5-0 both explicitly and implicitly follows logic models. For each of the elements of operational design and each of the JOPP steps, JP 5-0 explicitly lists the inputs to that element or step and the expected outputs. In both processes, many of the outputs of earlier steps or elements are then inputs to later steps. The overall presentation supports a logic model framework. For example, the emphasis in operational art on ends, ways, and means corresponds with logic model language: The ends are the outputs and outcomes, the ways are the activities, and the means are the inputs.

Logic Models Provide a Framework for Selecting and Prioritizing Measures

A logic model encapsulates a theory of change/logic of the effort and, done well, suggests things to measure.[7] Each layer in the logic model suggests clear measures. One might ask,

- Were all of the resources needed for the effort available? **(inputs)**
- Were all activities conducted as planned? On schedule? **(activities)**
- Did the activities produce what was intended? Did those products reach the desired audience? What proportion of that audience? **(outputs)**
- What proportion of the target audience engaged in the desired behavior? With what frequency? **(outcomes)**
- How much did the effort contribute to the overall campaign? **(impacts)**

These questions point directly to possible measures, and also help to prioritize. Not everything needs to be measured in great detail or particularly emphasized in data collection.[8] For example, the level of assessment data collection for inputs may be quick, simple, and holistic.

The benefit to measuring aspects of all of the different layers in the logic model is at its greatest when an effort is not working, or is not working as well as imagined. When the program does not produce all the expected outcomes and one wants to determine why, a logic model (or another articulation of a theory of change) really shines.

---

[6] Author interview with Maureen Taylor, April 4, 2013.

[7] Author interview with Christopher Nelson, February 18, 2013.

[8] Author interview with Ronald Rice, May 9, 2013.

Program Failure Versus Theory Failure

A program or effort does not produce the desired results (outcomes) for one of two fundamental reasons: either *program failure*, in which some aspect of the effort failed to produce the needed outputs, or *theory failure*, where the indicated outputs were produced but did not lead to the intended outcomes. Figure 5.2 illustrates the logic of program failure versus theory failure.

Logic model–based assessment can help identify which is the case and help initiate steps to improve the situation. If program failure is occurring, scrutiny of resources and activities can lead to process improvement and getting outputs on track. If the theory is flawed, it can be diagnosed, tweaked on the fly and experimented with, or replaced with an alternative theory (and supporting inputs, activities, and outputs).

Figure 5.2
Program Failure Versus Theory Failure
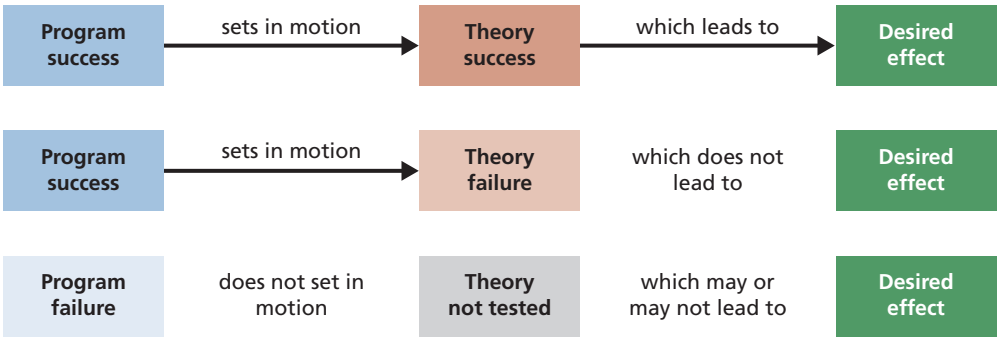
SOURCE: Thomas W. Valente, *Evaluating Health Promotion Programs*, Oxford, UK: Oxford University Press, 2002, p. 53, Figure 3.6. Used with permission.
RAND *RR809/2-5.2*

Constraints, Barriers, Disruptors, and Unintended Consequences

In addition to specifying inputs, activities, outputs, outcomes, and impacts, logic modeling (or other forms of articulating a theory of change/logic of the effort) provides an opportunity to think about things that might go wrong. Which assumptions are the most vulnerable? Which of the inputs are most likely to be late? Which of the activities might the adversary disrupt, or which activities are contingent on the weather? These things can be listed as part of the logic model and placed next to (or between) the nodes they might disrupt. For example, if local contractors might abscond with funds allocated for printing, or if the contractors are vulnerable to long power outages that can stop their presses, then these things could be noted between the relevant input and activity. If friendly force–caused collateral damage can prevent the translation of a short-term outcome into a long-term impact, it could be noted between outcomes and impacts.

Note that these disruptors can be anything outside the direct control of the program or effort.[9] For IIP efforts, this could include contextual factors (language, culture, history), exogenous shocks (natural disasters, economic crises, significant political action), actions by adversaries, actions by third parties in the information environment, and kinetic actions by friendly forces. The kinetic actions of a force send messages with far greater force than spoken or written messages.[10] If a picture is worth 1,000 words, then a JDAM (joint direct attack munition) is worth 10,000.[11]

If these potential disruptors can be conceived of as part of the logic modeling process, then, as needed, they can also be included in the measurement and data collection plan. The collection of such information can further facilitate the adjustment of situations involving apparent program or theory failure, or awareness that failure has come from an unanticipated and external source, and that neither the theory nor the program has actually failed—they have just been temporarily derailed by outside circumstances.

Barriers or disruptors do not necessarily completely disrupt processes (though some do), but all will at least slow down or diminish the rate of success. Perhaps they are best conceived like the "coefficient of friction" in physics. If desired levels of results (be they outputs or outcomes) are not being produced and an identified disruptor is measured as being present, adjustments can be made. These adjustments might simply be to put more of an input or activity in place (realizing that a certain amount is being lost to "friction"), or to identify some kind of workaround to minimize or remove the impact of the disruptor.

Further Reading

In this handbook:

Chapter Six discusses the development of measures for DoD IIP efforts, including types of measures and identifying constructs worth measuring.

In the accompanying desk reference:

Chapter Five offers a more comprehensive introduction to the concepts of logic models and theories of changes.

Building a Logic Model or Theory of Change

A theory of change/logic of the effort helps ensure that there are clear logical connections specified (either as assumptions or hypotheses, or a combination of both) between the activities of a program or effort and the objectives. Especially in the cognitive and behavioral realm, where shared understanding of such connections is lacking, explic-

---

[9]   Author interview with Christopher Nelson, February 18, 2013.

[10]   Author interview with Steve Booth-Butterfield, January 7, 2013.

[11]   Christopher Paul, *Strategic Communication: Origins, Concepts, and Current Debates*, Santa Barbara, Calif.: Praeger, 2011.

itly specifying the theory of change can be critical to both execution and assessment. A logic model is one way to articulate a theory of change. This section offers some concrete advice for the building or development of a program theory of change.

Various Frameworks, Templates, Techniques, and Tricks for Building Logic Models

Building a logic model is fundamentally about articulating the underlying logic of the program or effort.[12] To a certain degree, the framework of inputs to activities to outputs to outcomes to impacts is sufficient to begin to develop a logic model. Begin at the right, with SMART objectives, and work backward to the left.[13] What has to happen for those objectives to be met? What do you need to do to make those things happen? What resources do you need to do those things? A graphical depiction of this process of working backward appears in Figure 5.3.

Find and Fill Gaps in the Logic Model

Sometimes working backward from SMART objectives will result in more and more uncertainty at the levels of activities and inputs. In some situations (especially IIP situations), it is unclear what activities are most likely to produce the outputs needed to reach desired outcomes. When this occurs, additional information is needed.

**Figure 5.3
Working Backward to Articulate a Theory of Change**
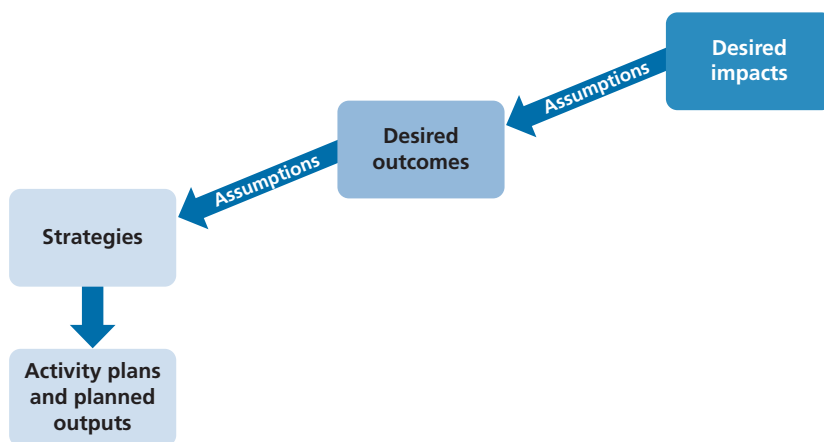
SOURCE: NATO, Joint Analysis and Lessons Learned Centre, 2013, p. 8, Figure 2.
RAND *RR809/2-5.3*

---

[12] There are a number of specific frameworks, worksheets, and guidebooks that can help with articulating a logic model or theory of change. We found two to be particularly relevant: North Atlantic Treaty Organization (NATO), Joint Analysis and Lessons Learned Centre, *A Framework for the Strategic Planning and Evaluation of Public Diplomacy*, Lisbon, Portugal, June 2013; and U.S. Agency for International Development, "Logical Framework Template: Basic," web page, undated.

[13] NATO, Joint Analysis and Lessons Learned Centre, 2013.

One approach to resolving uncertainty about the best activities to achieve desired outcomes is formative research. Formative research can help identify the mediating factors and test which kinds of messages or activities have the most influence on those factors; after such formative research, one is left not only with a thoughtfully articulated logic model but also with one that is at least partially validated. Formative research for this purpose might involve quick field experiments, pilot tests of draft products, focus groups with SMEs, or even a review of historical cases.[14] Methods and approaches to formative research are discussed further in Chapter Eight.

Another way to find and fill gaps in a logic model is based on operational experiences. The after-action review process is dedicated specifically to learning from both success and failure. As much as the tradition of the after-action review warrants praise for its ability to extract lessons learned from successful and unsuccessful campaigns, the approach has a major shortcoming that makes it an imperfect analogy for the assessment process: It is retrospective and timed in a way that makes it difficult for campaigns that are going to fail to do so quickly. On the other hand, JP 5-0 describes operational design as an iterative process, a process that can iterate not just during initial planning but also during operations as assumptions and plans are forced to change. Operational design also advocates continuous learning and adaptation, and well-structured assessment can support that. As we advocate in Chapter Two, fail fast! If a logic model contains uncertain assumptions, plan not only to carefully measure things associated with those assumptions but also to measure them early and often. If faulty assumptions are exposed quickly, this information can feed back into a new iteration of operational design, producing a revised logic model and operational approach.

Start Big and Prune, or Start Small and Grow

There is at least as much art as science to achieving the right level of detail in a logic model or theory of change. For example, a theory of change might begin as something quite simple: *Training and arming local security guards will lead to increased stability.* While this gets at the kernel of the idea, it is not particularly complete as a logic model. It specifies an outcome (increased stability) and some outputs (trained local security guards and armed local security guards), and further implies inputs and activities (the items needed to train and arm guards), but it does not make a clear, logical connection between the outputs and the outcome. Stopping with that minimal logic model could lead to assessments that would only measure the activity and the outcome. However, such assessments would leave a huge assumptive gap. If training and arming go well but stability does not increase, assessors will have no idea why. To begin to expand on a simple theory of change, ask the questions, "Why? How might A lead to B?" (In this case, how do you think training and arming will lead to stability?) A thoughtful answer to this question usually leads one to add another node to the theory of change,

---

[14] Author interview with Christopher Nelson, February 18, 2013.

or an additional specification to the logic model. If needed, the question can be asked again relative to this new node until the theory of change is sufficiently articulated.

How do you know when the theory of change is sufficiently articulated? There is no hard-and-fast rule. Too many nodes, too much detail, and you end up with something like the infamous spaghetti diagram of Afghan stability and counterinsurgency dynamics.[15] Add too few nodes and you end up with something too simple that leaves too many assumptive gaps. If an added node invokes thoughts such as, "Well, that's pretty obvious," perhaps it is overly detailed.

Elicit an Implicit Theory of Change

As noted, one challenge that can come up in logic modeling is when the inputs, activities, outputs, and outcomes are all clear, but it is not clear how the outputs are supposed to lead to the desired outcomes. This is a situation with an implicit logic of the effort, and the goal then becomes making it explicit. Faced with this situation, assessors can start by asking why and how questions (as suggested in the previous section), but it is possible that they will not be able to come up with satisfactory answers. Presumably, those engaged in the planning and execution of a program or activity have some idea why they do the things they do. Engaging stakeholders may quickly reveal missing connections in a theory of change. However, it is also possible that while stakeholders intuit how their actions connect to desired outcomes, they have a hard time articulating it. In such a case, the theory of change remains implicit, but working with stakeholders can still bring it to light. Begin with some specific program element and ask, "Why are you doing that?"[16] Break it down, walk through activities, and try to expose the internal logic of the effort or its shared understandings.

Updating the Theory of Change

Fortunately, if an initial theory of change is not sufficiently detailed in the right places or does not fit well in a specific operating context, iterative assessments will reveal where additional detail is required. Following the example discussion of a logic model for increasing stability by training and arming local security guards, imagine a situation in which measures show real increases in security (reduced violence and casualties, seasonally adjusted) but measures of perception of security (from surveys, focus groups, observed market attendance) do not correspond. If planners are not willing to give up on the assumption that improvements in security lead to improvements in perceptions of security, they can speculate and add another node, or they can do some quick data

---

[15] In 2009, GEN Stanley McChrystal, then commander of U.S. and NATO forces in Afghanistan, received a PowerPoint slide meant to convey the complexity of the coalition military strategy for counterinsurgency and stability operations in that country. The slide prompted two strains of commentary: one declaring that the Afghanistan strategy had gotten out of hand and another declaring that the military's use of PowerPoint had gotten out of hand. We revisit both these points in Chapter Eleven, on the presentation and uses of assessment.

[16] Rossi, Lipsey, and Freeman, 2004, p. 148.

collection, getting a hypothesis from personnel operating in the area or from a local focus group. Perhaps the missing node is *awareness* of the changing security situation. If preliminary information confirms this as a plausible gap, then it also indicates the need for a new activity in addition to a new node: some kind of effort to increase awareness of changes in the security situation.

Improvements to the theory of change improve assessments, and they can *also* improve operations. Further, articulating a theory of change during planning allows activities to begin with some questionable assumptions in place—and with the confidence that they will be either validated by assessment or revised. Theory of change–based assessment supports learning and adapting in operations. (Again, as we advocate in Chapter Two, fail fast.)

Validating Logic Models

Logic models should be validated. Sometimes IIP programs or efforts are predicated on incorrect assumptions. Sometimes IIP efforts are based on a thoughtful foundation derived from existing psychological research, but that foundation is not applicable in the given cultural context. As noted previously, one way to validate a logic model is to execute based on it, revise it through trial and error, and declare it valid when it finally works. The summative evaluation for a successful effort or program validates the program's logic model.[17]

Logic models can also be validated in other ways. One such approach is similar to the formative research recommended earlier for building a logic model: some sort of SME engagement. If a preliminary logic model survives scrutiny by a panel of both influence and contextual experts, then it is likely to last longer and with fewer subsequent changes than a logic model not validated in this way. In JOPP, this could be part of COA analysis and war-gaming, though the logic model may require input from SMEs outside the standard staff.

Further Reading

In this handbook:

Chapter Two, in the section "Effective Assessment Requires a Theory of Change or Logic of the Effort Connecting Activities to Objectives," articulates the connection between a theory of change and best assessment practices.

Chapter Six discusses the development of measures for DoD IIP efforts, including types of measures and identifying constructs worth measuring.

In the accompanying desk reference:

Chapter Five offers a more comprehensive introduction to the concepts of logic models and theories of change.

---

[17] Author interview with Christopher Nelson, February 18, 2013.

Key Takeaways

- Specifying a theory of change involves identifying overall objectives, as well as the inputs, outputs, and processes necessary to achieve those objectives, and describing the logic that underpins it all (an explanation of how the proposed actions will lead to the desired outcomes). A logic model is one structure for presenting a theory of change.
- A program's theory of change contains assumptions about how the world works and what kinds of activities will lead to desired goals and why. Assessment can help distinguish between theory failure (one or more of the assumptions is wrong) and program failure (the program is not being executed properly); assessment can also help identify ways to correct either of these failings.
- In addition to describing the logical connections between activities and objectives, a good theory of change should include possible barriers, disruptors, threats, or alternative assumptions. If things that might divert progress and prevent objectives from being achieved are identified at the outset, they can be included in the assessment process.
- Logic models often require revision when exposed to reality. Iteration and evolution are important to (and expected of) theories of change.
- Logic models should be validated. This can be accomplished through SME engagement, through other research efforts, or through trial and error as part of assessment within a program of activities.
- When the program does not produce all the expected outcomes and one wants to determine why, a logic model (or other articulation of a theory of change) really shines.

Developing Measures for DoD IIP Efforts

Here, we address the processes and principles that govern the development of valid, reliable, feasible, and useful measures that can be used to assess the effectiveness of IIP activities and campaigns. The development of measures is decomposed into two broad processes:

1. deciding what constructs are essential to measure
2. operationally defining the measures.

Ideally, an assessment should include a measure to gauge every cause-and-effect relationship specified in the program logic model. DoD assessment doctrine emphasizes the distinction between MOPs and measures of effectiveness (MOEs). In IIP evaluation, MOEs are typically associated with attitudinal and behavioral changes at the individual and group levels. Whether attitudinal change constitutes an effect is controversial, which demonstrates a limitation to the MOP-versus-MOE construct.

While appreciating the conceptual differences between measure types can be valuable, assessment reports should avoid being overly concerned with the difference between MOPs and MOEs, because this focus is overly narrow and potentially distracting. In reality, there is a spectrum of measure types, and the MOE-MOP dichotomy can mislead evaluators into thinking that there are only two relevant measures. At worst, premature conclusions made on the basis of a single MOE can lead to the termination of an otherwise promising effort.

Further Reading

In this handbook:

Chapter Four, in the sections "Behavioral Versus Attitudinal Objectives" and "Intermediate Versus Long-Term Objectives," discusses distinctions between different types of objectives.
Chapter Five, in the section "Program Failure Versus Theory Failure," addresses points of failure.

In the accompanying desk reference:

Chapter Six, in the section "Hierarchy of Terms and Concepts: From Constructs to Measures to Data," clarifies the terms and concepts of measure development. Also, the section "Types of Measures" explores in greater detail the pitfalls of the distinction between MOPs and MOEs, including its articulation in JP 5-0.
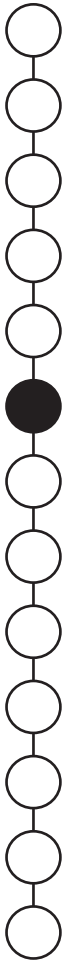
Identifying the Constructs Worth Measuring: The Relationship Between the Logic Model and Measure Selection

Separating what is important to measure from what is less important "is what measure development is all about."[1] The program logic model provides the framework for selecting the constructs that are worth measuring, but evaluators should not assume that all important measures will simply "fall into their laps" in the course of planning. As Christopher Nelson points out, goals and objectives can be unclear or unmeasurable, and program managers often disagree on the ultimate goal that a program is designed to serve.[2] Moreover, it is too costly to measure every cause-and-effect relationship and mediating variable within the system that ties program inputs to outputs to outcomes.

The importance of measuring something, or the information value of a measure, is a function of uncertainty about its value and the costs of being wrong. When identifying constructs worth measuring, assessors should therefore give priority to "load-bearing" and vulnerable cause-and-effect relationships in the logic model. These can be identified by drawing on IIP theories, empirical research, expert elicitation, and rigorous evaluations of similar programs implemented in the past.[3] Moreover, the information value of a measure takes precedence over its validity and reliability. Even the most valid and reliable measurement instruments cannot improve the value of the measure if it is measuring a construct that is irrelevant to assessment stakeholders and the decision they need to make. Assessors should therefore try to measure every truly important variable even if the measurement instrument has weak validity. Douglas Hubbard emphasizes this point in *How to Measure Anything*: "If you are betting a lot of money on the outcome of a variable that has a lot of uncertainty, then even a marginal reduction in your uncertainty has a computable monetary value."[4]

Attributes of Good Measures

The quality of a measure is typically evaluated on the basis of its *validity*, *reliability*, *feasibility*, and *utility*:

- *Validity* is the correspondence between the measure and the construct—or freedom from systemic error (bias).
- *Reliability* is the degree of consistency in measurement—or freedom from random error (e.g., signal to noise).

---

[1]   Author interview with Christopher Nelson, February 18, 2013.

[2]   Author interview with Christopher Nelson, February 18, 2013.

[3]   Author interview with Christopher Nelson, February 18, 2013.

[4]   Hubbard, 2010, p. 36.

Box 6.1
Where to Begin? Measuring Baselines and Variables

Before any IIP intervention, there exists a *prior state* that characterizes the people, their attitudes, their community, the security environment, the economy, and so on. This serves as the baseline with which evaluation measurements are compared. This prior state will also include constraints that need to be considered and could affect the success of an assessment if they're not. These constraints are not limited to characteristics of the local environment (such as security concerns). They could also include the need to work around another operation, such as a counterinsurgency operation or a kinetic operation in the same area.

If an operation kills innocent civilians, for example, there is very little that a communication campaign can do to shape the information environment to counteract that. On the other hand, kinetic and information operations should be mutually supportive. It is important to control for noncommunicative aspects of the campaign to identify the unique contributions of the communication campaign as well as the extent to which both components are mutually supportive.[a]

A key aspect in evaluating a complex campaign is the need to consider, measure, and assess the effect of these factors to explain why certain outputs occurred and others did not.[b] This is part of the iterative operational design process prescribed in JP 5-0, especially the imperative to understand the operational environment.

How often do prior states and system variables need to be measured? The answer depends on the rate at which the variables are expected to change over time. Some things are very slow to change and therefore typically only need to be measured once (e.g., the presence of a health care clinic). But variables that change frequently—such as kinetic operations or economic conditions—should be measured often, at intervals sufficient to capture relevant change.[c]

[a] Author interview with Steve Booth-Butterfield, January 7, 2013.

[b] Ronald E. Rice and Dennis R. Foote, "A Systems-Based Evaluation Planning Model for Health Communication Campaigns in Developing Countries," in Ronald E. Rice and Charles K. Atkin, eds., *Public Communication Campaigns*, 4th ed., Thousand Oaks, Calif.: Sage Publications, 2013.

[c] Author interview with Ronald Rice, May 9, 2013.

- *Feasibility* is the extent to which data can actually be generated to populate the measure with a reasonable level of effort.
- *Utility* is the usefulness of the measure to assessment end users and stakeholders.[5]

Validity and reliability represent the two types of measurement error. There is tension between the feasibility of a measure and its utility. Often, what is important or useful to measure cannot be easily observed. It is important to first identify the measures with the highest information value and subsequently determine what is feasible among those worth measuring.

Further Reading

In this handbook:

Chapter Three provides more background on the utility of measures in the context of users of assessment results.

---

[5]   Author interview with Christopher Nelson, February 18, 2013.

Chapter Four explains how the quality of measures depends to a great degree on the quality of the objectives articulated during the planning phase. The same principles guide the development of objectives, logic models, and the measurement system.

Chapter Five discusses the attributes of logic models that facilitate effective measurement and assessment.

In the accompanying desk reference:

Chapter Six, in the section "Identifying the Constructs Worth Measuring: The Relationship Between the Logic Model and Measure Selection," offers a more detailed discussion of determining what to measure. The following sections also break out and address in-depth the attributes of good measures reviewed here:

- "Assessing Validity: Are You Measuring What You Intend to Measure?"
- "Assessing Reliability: If You Measure It over Again, Will the Value Change?"
- "Assessing Feasibility: Can Data Be Collected for the Measure with a Reasonable Level of Effort?"
- "Assessing Utility: What is the Information Value of the Measure?"
- "Feasibility Versus Utility: Are You Measuring What Is Easy to Observe or Measuring What Matters?"

Developing Measures: Advice for Practitioners

Keep a record of validated and potential IIP measures and indicators.
Although a repository would be ideal, a more practical solution for practitioners could be to keep records on where measures have been used before, how well they worked, and the evidence that supports them. It might be useful to also keep records of invalid measures and indicators to avoid using them again.

Tie each influence objective to several specific measures.
Some measures will have insufficient or unreliable data and need as much support as possible. Suppose your goal is to reduce the influence of a particular mullah. Your measures could assess (1) the population's self-reported impressions of him; (2) attendance at his mosque; and (3) how often he is mentioned in communications from various organizations or the press.[6]

Avoid "metric bloat" or "promiscuous" measure collection.
Having too many measures per objective can complicate analysis and the interpretation of results.[7] If the number of measures is becoming unmanageable, discard the lower-performing ones. It is also worth noting that measuring the same outcome twice does not satisfy two layers of the assessment scheme. For example, "Reductions in the number of attacks and incidents will lead to increased security" almost sounds sensible, but this is what it really says: "Increases in security will lead to increased security."

---

[6]   Author interview with Anthony Pratkanis, March 26, 2013.

[7]   William P. Upshur, Jonathan W. Roginski, and David J. Kilcullen, "Recognizing Systems in Afghanistan: Lessons Learned and New Approaches to Operational Assessments," *Prism*, Vol. 3, No. 3, 2012, p. 91; Stephen Downes-Martin, "Operations Assessment in Afghanistan Is Broken: What Is to Be Done?" *Naval War College Review*, Vol. 64, No. 4, Fall 2011, p. 108.

Express numeric measures in the form of a ratio so that progress from the baseline to future states can be easily determined.

In this formulation, the baseline value is the denominator and changes due to the IIP activity are reflected in the numerator.[8]

Avoid the temptation to collect data only on indicators of success.

Measures or indicators should be defined or scaled so that they capture failure or regression as well as success.[9] The measurement system should also be flexible enough to capture unintended consequences.[10] When things are going well, it may be tempting to only measure outcomes, but assessment is at its best when things are *not* going well. Measuring intermediate nodes in a theory of change can help determine *why*. As mentioned in Chapter Five, this is when a logic model (or other articulation of a theory of change) really shines.

Avoid perverse incentives.

A perverse incentive is an incentive (usually an unintended one) that rewards an undesirable result. Measures of exposure are particularly susceptible to perverse incentives.[11] A recent State Department Inspector General's report accused the Bureau of International Information Programs of "buying likes" on Facebook as a way to improve the perceived reach of a program.[12] Such a strategy may increase awareness, but it will not tell you anything about a program's impact.

Avoid measures that are easily manipulated.

Past examples of manipulated or "captured" metrics in counterinsurgency environments have included exaggerated reports of the operational readiness of host-nation forces or of enemy casualties and reduced reporting of civilian casualties.[13] Careful data collection, in addition to careful measure selection, can help mitigate this risk.

Further Reading

In the accompanying desk reference:

Chapter Six, in the section "Constructing the Measures: Techniques and Best Practices for Operationally Defining the Constructs Worth Measuring," expands on the advice presented here.

---

[8] The Initiatives Group, 2013.

[9] Author interview with Steve Booth-Butterfield, January 7, 2013.

[10] Author interview with James Pamment, May 24, 2013.

[11] Author interview with Craig Hayden, June 21, 2013.

[12] Office of the Inspector General, U.S. Department of State, *Inspection of the Bureau of International Information Programs*, May 2013; Craig Hayden, "Another Perspective on IIP Social Media Strategy," *Intermap*, July 23, 2013.

[13] Dave LaRivee, *Best Practices Guide for Conducting Assessments in Counterinsurgencies*, Washington, D.C.: U.S. Air Force Academy, December 2011, p. 18.

Key Takeaways

- The quality of measures depends on the quality of the objectives enumerated in the program's logic model.
- The importance of measuring something, or the information value of a measure, is determined by the amount of uncertainty about its value and the costs of being wrong. Assessors should therefore give priority to "load bearing" or vulnerable processes. These elements can be identified through IIP theories, empirical research, expert elicitation, and evaluations of similar campaigns implemented in the past.[14]
- Good measures are valid, reliable, feasible, and useful.
- There is tension between the feasibility of a measure and its utility. Often, what is important or useful to measure cannot be easily observed. Assessors should first identify the measures with the highest information value and subsequently determine what is feasible among those worth measuring.
- Engage in best practices for measure development, including keeping records of what has been successful and not successful, tying objectives to several specific measures, avoiding "metric bloat," expressing numeric measures in the form of a ratio, avoiding the temptation to collect data only on indicators of success, and avoiding perverse incentives and measures that are easily manipulated.
- Many measures will only be useful when things are not going well, but they may be essential to diagnosing and correcting a problem.

---

[14] Author interview with Christopher Nelson, February 18, 2013.

Designing and Implementing Assessments

The *design* of an assessment or evaluation is the plan that describes the research activities that will answer the questions motivating the evaluation. The design determines the way in which the evaluation can (or cannot) make causal inferences regarding the outputs, outcomes, or impacts of the intervention. Design-related decisions govern the *structure* of data collection (i.e., the number, timing, and type of data measurements), rather than the *methods* by which data are collected. There are three broad types of evaluation design:

- experimental (control with random selection)
- quasi-experimental (control without random selection)
- nonexperimental or observational studies (no control).

Practitioners should already be familiar with a range of potential evaluation designs and their strengths and weaknesses so that they can design the best and most appropriate evaluation given stakeholders' needs, populations affected, and available resources.[1] Therefore, we do not spend a great deal of time on the topic in this handbook.

Criteria for High-Quality Evaluation Design: Feasibility, Validity, and Utility

How should evaluators choose among possible evaluation designs? This section proposes that the best designs are feasible, valid, and useful. However, there are tensions and trade-offs inherent in pursuing each of those objectives. It is important to select the strongest evaluation design, in terms of internal and external validity, among those designs that are useful and feasible with allocated resources.[2] However, the most rigorous design varies with the importance and intended use of the results. Resources should therefore be allocated according to the importance of potential outcomes. In a budget-constrained environment, evaluations are *simultaneously more important and less affordable*. To allow room for

---

[1]   Valente, 2002, pp. 87–88.

[2]   Valente, 2002, pp. 89–90.

more assessments within budget constraints, there needs to be a mechanism for quick, cheap, and "good enough" assessments.

Designing Feasible Assessments

Acknowledging the importance of constructing the best and most valid evaluation possible given the available resources, Thomas Valente states that the first requirement of evaluation design "is that it be practical, which often prevents the use of the best design that might be theoretically possible."[3] Time, resources, and ethical or practical concerns with carrying out randomized experiments all constrain feasibility.

To gauge the feasibility of a new, resource-intensive evaluation design, IIP evaluators should consider using *pilot evaluations*. Pilot evaluations test the evaluation design on a much smaller scale than ultimately envisioned by either studying the effectiveness of a small effort or focusing on a subset of the target audience. Time permitting, DoD IIP efforts should include both pilot tests of the effort's activities and pilot tests of the evaluation design. Such limited-scope formative efforts can ensure that money for the full-scale efforts is well spent.

Designing Valid Assessments

Designing feasible evaluations is in tension with designing valid ones. Validity represents the extent to which a design or a measure is accurate or free from systemic bias. *Internal validity* is the extent to which the design supports the kinds of causal inferences or causal conclusions that need to be made within the evaluation. *External validity* (also known as *generalizability* or *ecological validity*) is the extent to which design is able to support inference (e.g., generalize) about the larger population of interest.

In the DoD context, the contribution of the IIP effort often cannot be separated from "background noise" and operational, tactical, and strategic factors.[4] Adding to the complexity is the challenge associated with isolating the contribution of influence tactics within the broader context of a military campaign. The most-valid evaluations are those that include the most-effective controls against those factors. However, such designs will be more complex and therefore (typically) more resource intensive.

There is often a trade-off between external and internal validity. Designs with the highest internal validity often have weak ecological validity, because the "laboratory-like" conditions required to control for the threats to internal validity do not appropriately reflect conditions in which the focal audience would interact with the program "in the wild" or under generalizable circumstances.[5] Likewise, field experiments taking place "in the wild" have the highest ecological validity but are the hardest to control for threats to internal validity.

---

[3] Valente, 2002, p. 88.

[4] David C. Becker and Robert Grossman-Vermaas, "Metrics for the Haiti Stabilization Initiative," *Prism*, Vol. 2, No. 2, March 2011.

[5] Author interview with Marie-Louise Mares, May 17, 2013.

Designing Useful Assessments

As emphasized throughout this handbook, assessment is a decision-support tool. The way in which the assessment will be used has significant implications for an assessment's design. Assessment design, processes, and degree of academic rigor and formality should be tailored to the assessment end users and stakeholders. Field commanders and congressional leaders will have different sets of questions.[6] Part of successful assessment design is balancing stakeholder needs with feasibility and rigor.

To design a useful evaluation, evaluators must first understand the assessment audience (users and stakeholders) and the decisions it will inform (assessment *uses*). *End users* are those with formal or institutional responsibility and authority over the program and have an active interest in the evaluation. In the IO context, program managers, military leadership, and Congress represent potential end users, depending on the level of evaluation. *Stakeholders* include a broader set of "right-to-know" audiences that have a more passive interest in the evaluation. Stakeholders could include the target audience, media, and internal program management and staff.[7]

As noted in Chapter Three, there are three primary uses for assessment: planning, improvement, and accountability. These categories roughly correspond to the three types, or stages, of evaluation: formative, process, and summative. Accountability-oriented evaluations will tend to target end users outside DoD. Improvement-oriented evaluations have end users who are internal to the program.

To get a better idea of users and uses, it may be helpful to create a matrix similar to the one shown in Table 7.1, which maps each assessment user to an assessment use.[8] The matrix can be color-coded to show immediate, medium-term, and long-term needs.

Further Reading
In this handbook:

Chapter Two, in the section "Assessment Requires Resources," touches on the notion that not all assessments need the same level of depth or quality.

Chapter Three provides more detail on the primary users and uses of DoD IIP assessment results, including how formative and process evaluation support improvement-oriented assessment and how summative assessment supports accountability-oriented assessment.

Chapter Six, in the section "Attributes of Good Measures," discusses these attributes as they pertain to measures.

In the accompanying desk reference:

Chapter Seven provides more detail on the extent to which various study designs control against threats to internal validity (see, especially, Table 7.2). That chapter also includes an example of a populated users-uses matrix (Table 7.5).

Chapter Eleven, in the section "Evaluating Evaluations: Meta-Analysis," addresses the process of assessing assessments to these and other standards.

---

[6] Author interview with Monroe Price, July 19, 2013.

[7] Author interview with Christopher Nelson, February 18, 2013.

[8] Author interview with Christopher Nelson, February 18, 2013.

Table 7.1
Uses and Users Matrix Template

		Likely Uses		
		Accountability	Improvement	Combined/Other
Likely Users	End users			
	Stakeholders			
	Others			

Formative Evaluation Design

Formative evaluation is the preintervention research that helps to shape the campaign logic model and execution. Formative evaluation can define the scope of the problem, identify possible campaign strategies, provide information about the target audience, determine what messages work best and how they should be framed, determine the most-credible messengers, and identify the factors that can help or hinder the campaigns.[9]

Formative evaluation design can range from observational studies using focus groups, interviews, atmospherics, or baseline surveys to laboratory experiments for testing the efficacy of messages and media. To inform decisionmaking, formative research must be turned around quickly. It should also feed back into the logic model development and refinement process.

Process Evaluation Design

Process evaluation serves several purposes and is underutilized. Process research can document implementation, guide program adjustments mid-implementation, identify whether the necessary conditions for impact took place, identify the causes of failure (see "Program Failure Versus Theory Failure" in Chapter Five), identify threats to internal validity (such as contamination or interference from other campaigns), and generate information necessary for replicating and improving the program or campaign.

Summative Evaluation Design

Summative evaluations consist of postintervention research designed to determine the outcomes that can be attributed or tied to the IIP intervention or campaign. Determining causality—or the extent to which one or more influence activities contributed to or

---

[9] Julia Coffman, *Public Communication Campaign Evaluation*, Washington, D.C.: Communications Consortium Media Center, May 2002.

was responsible for a change in knowledge, attitudes, or behaviors—is a chief goal of summative IIP evaluation.[10] Summative evaluation designs can be classified as *experimental*, *quasi-experimental*, or *nonexperimental*.

In experimental designs, subjects are randomly assigned to treatment and control conditions and are observed, at minimum, after treatment. Experimental designs have the highest internal validity and therefore the strongest basis for causal inference. Quasi-experimental designs or *natural experiments*, such as longitudinal or cross-sectional *exposed versus unexposed* studies, are similar to experimental designs except that the researchers cannot randomly assign subjects to treatment or control groups. Quasi-experimental evaluation designs can be mixed method, incorporating qualitative components. Quasi-experimental designs have lower internal validity than experimental designs but are often much more practical and cost-effective. Nonexperimental studies do not have a control and therefore have limited to no ability to make causal claims regarding the contribution of the program to outcomes, but they can nonetheless be useful to gather information on perceptions of the campaign.

Within those broad categories there are many design variations. The following were among those reviewed for this research: field experiments and randomized controlled trials (experimental); variations on exposed-versus-unexposed designs, split or "A/B" testing, the "bellwether" method, and longitudinal designs (quasi-experimental); and frame evaluation research and case studies (nonexperimental). Organizations with effective research cultures often use several designs.

Further Reading

In this handbook:

Chapter Three offers an introduction to formative, process, and summative evaluation, including additional background on characteristics and the hierarchy of evaluation.

Chapter Five, in the section "Program Failure Versus Theory Failure," discusses possible reasons for failure, which process evaluation can help determine.

In the accompanying desk reference:

Chapter Seven, in the section "Experimental Designs in IIP Evaluation," discusses the appropriateness of experimental designs for IIP evaluation and the special case of survey experiments. That chapter also reviews quasi-experimental and nonexperimental designs in greater detail, including examples of these designs in practice drawn from across the sectors examined in this research; see the following sections:
  • "Quasi-Experimental Designs in IIP Evaluation"
  • "Nonexperimental Designs"

The Best Evaluations Draw from a Compendium of Studies with Multiple Designs and Approaches

Each design has strengths and weaknesses that vary by environment and circumstance. No single design will be appropriate for all campaigns. And, independent of feasibility, no single design will present a full picture of effectiveness. Thus, the most valid conclu-

---

[10] Valente, 2002, p. 89; author interview with Kavita Abraham Dowsing, May 23, 2013.

sions about program effects are those that are based on results from multiple studies using different designs. Even if they are feasible, using the same approaches over and over leads only to a partial answer, which can be a mistaken answer, "so the best way to do research is to approach it from multiple angles—surveys, some experimental work, in-depth interviews, and observational work."[11]

Steve Booth-Butterfield makes that case that triangulation is particularly important in IIP evaluation due to the challenges with data availability and quality.[12] Because there are limitations to each approach, IIP evaluators should look at all evidence from as many different angles that are reasonable, rational, empirical, and feasible and see whether the evidence is trending in the same direction. While it is relatively easy to identify weaknesses with any single measure, when a collection of measures across different methods is suggesting the same general trend, you can have much more confidence in your conclusions.

Further Reading

In the accompanying desk reference:

Chapter Seven, in the sections "The Best Evaluations Draw from a Compendium of Studies with Multiple Designs and Approaches" and "The Importance of Baseline Data to Summative Evaluations," offers an expanded discussion of triangulation and the importance of baseline data, respectively.

Box 7.1
The Challenge of Determining Causality in IIP Evaluation

There are many daunting challenges to establishing causality in IIP evaluations. But despite these difficulties, it is not impossible to obtain reasonable estimates of causal effects. A DoD MISO practitioner commented that much of the concern over causality is driven by a lack of awareness of alternatives to true experimental design.[a] In *Data-Driven Marketing: The 15 Metrics Everyone in Marketing Should Know*, Mark Jeffrey responds to the objection that there are too many factors to isolate cause and effect: "The idea is conceptually simple: conduct a small experiment, isolating as many variables as possible, to see what works and what does not."[b]

Ultimately, there are a number of designs that can lead to assessments of DoD IIP activities with high internal validity and allow strong causal claims. These designs tend to be more resource intensive, and they require an unambiguous commitment to some kind of experimental or quasi-experimental structure in program delivery and assessment. This, then, turns back to the matter of feasibility. If you want to be able to make causal claims, are you willing to put forward the time and effort necessary to make that possible?

While experimental or quasi-experimental designs are often comparatively resource intensive, many quasi-experimental designs are more feasible in the defense context than many planners might think. A functional quasi-experimental design may simply require a delay in delivery of all or part of a program's materials and outcome measurements at a few additional time points. Quasi-experiments are not as rigorous as randomized controlled experiments, but they still provide strong grounds from which to assert causation—sufficient for many assessment processes.

[a] Author interview on a not-for-attribution basis, July 30, 2013.

[b] Mark Jeffery, *Data-Driven Marketing: The 15 Metrics Everyone in Marketing Should Know*, Hoboken, N.J.: John Wiley and Sons, 2010.

---

[11] Author interview with Devra Moehler, May 31, 2013.

[12] Author interview with Steve Booth-Butterfield, January 7, 2013.

Key Takeaways

- The best designs are valid, generalizable, practical, and useful. However, there are tensions and trade-offs inherent in pursuing each of those objectives. Evaluators should select the strongest evaluation design, using a methodological perspective, from among those designs that are feasible with a reasonable level of effort and resources.
- Assessment design, processes, and level of rigor and formality should be tailored to the assessment end users and stakeholders. Academic rigor must be balanced with stakeholder needs, appetite for research, and cost considerations.
- Formative research must be turned around quickly to inform decisionmaking.
- Internal validity is the extent to which the design of the evaluation supports the causal inferences it purports to make. Internal validity is limited by confounding variables, selection bias, maturation, history, instrumentation, attrition, and regression toward the mean.
- Threats to internal validity are controlled by design choices. Broadly, designs can be classified as experimental (random assignment with a control group), quasi-experimental (comparison group without random assignment), or nonexperimental (no comparison group). The more controlled the design, the higher the internal validity. Thus, the relative value of experimental research depends on the importance of making causal inference.
- Determining causality in the defense IIP context is not as difficult as you might think. When determining causality is important, quasi-experimental designs will often be the best (balancing practicality, rigor, and utility) design option available.
- To balance the strengths and weaknesses across different designs, the best evaluations draw from a compendium of studies with multiple designs and methods that converge on key results. Implementing this approach requires a single person or group "at the top" with responsibility for triangulating the disparate approaches.

Formative and Qualitative Research Methods for DoD IIP Efforts

While formative and qualitative research often overlap, they are by no means completely equivalent. Formative evaluations can use quantitative methods, and qualitative methods can inform evaluations conducted in each of the three phases (formative, process, and summative).

Formative research methods are varied. Classical methods include focus groups and in-depth interviews. Increasingly, researchers are relying more on quantitative approaches, such as content analysis and laboratory experiments, to test the cognitive effects of messages and products. Less traditional qualitative methods encountered in our research include community assessments, photojournalism, and temperature maps.[1]

The Importance and Role of Formative Research

Several of the SMEs interviewed stressed the importance of formative research and argued that it is systemically undervalued, especially in periods of budgetary cutbacks. However, an up-front investment in formative research typically saves costs in the long run because it increases the likelihood that the program will be effective, reduces expenses associated with program implementation, and saves costs during both the process and summative evaluation phases.[2] By demonstrating the likely effects of the effort on targeted audiences, formative research allows practitioners to have greater confidence in their conclusions about the expected effects of an effort. If an effort has been validated as having a certain effect, campaign effectiveness will then depend principally on the extent of exposure.[3] Likewise, if summative research shows a lack of outcomes, evaluators can more easily isolate the source of program failure if they conducted sound formative research.

---

[1] Author interview with Kavita Abraham Dowsing, May 23, 2013.

[2] Author interview with Thomas Valente, June 18, 2013; author interview with Charlotte Cole, May 29, 2013.

[3] Author interview with Mark Helmke, May 6, 2013.

Identifying the Audience and Characterizing the Information Environment

The first component of formative research is to determine the characteristics of the target audience and information environment (IE) that shape views and behaviors. The first step in the *Joint Information Operations Assessment Framework*, for example, is to characterize the IE, including the "cognitive, informational, and physical domains," to inform campaign planning.[4] "Understand the operational environment" is a key imperative of operational design and is a predicate for mission analysis in JOPP, according to JP 5-0. Other guidance may refer to this process as the "needs assessment" or as measuring the "system of influence" that the intervention is operating within. This section explores three key, interrelated analytic tasks associated with this phase: audience segmentation, social network analysis, and target audience analysis.

Audience Segmentation

Audiences are not homogeneous groups. Audience segmentation techniques help planners understand how different messages resonate with different segments of the population.[5] IIP interventions should differentiate populations into segments of people that share "needs, wants, lifestyles, behaviors and values" that make them likely to respond similarly to an intervention.[6]

When it comes to message receptiveness, demographic segmentation often poorly reflects diversity within a population. Better approaches segment the audience along psychographic variables and their demographic correlations rather than on demographic variables alone.[7] Rather than assuming that people of a similar race, gender, or age share similar values, planners should segment the audience according to what is important to them and subsequently determine whether those values correspond to demographic categories.

For awareness campaigns, some social marketing experts suggest that audiences should be segmented by self-rated prior knowledge. Andrea Stanaland and Linda Golden have observed that people with higher self-rated knowledge are not message receptive, presumably because they do not feel a need for additional information. In this sense, self-rated knowledge may diminish the motivation to process new information, adversely affecting message receptivity.[8]

---

[4]   Joint Information Operations Warfare Center, *Joint Information Operations Assessment Framework*, October 1, 2012, pp. 11–12.

[5]   Author interview with Gerry Power, April 10, 2013.

[6]   Sonya Grier and Carol A. Bryant, "Social Marketing in Public Health," *Annual Review of Public Health*, Vol. 26, 2005, p. 322.

[7]   Author interview with Gerry Power, April 10, 2013.

[8]   Andrea J. S. Stanaland and Linda L. Golden, "Consumer Receptivity to Social Marketing Information: The Role of Self-Rated Knowledge and Knowledge Accuracy," *Academy of Marketing Studies Journal*, Vol. 13, No. 2, 2009, p. 32.

Social Network Analysis

Network analysis, also known as *social network analysis*, can improve campaign strategy and targeting by identifying key influencers and opinion leaders. Opinion leaders typically have greater exposure to messages and are more likely to exercise informal influence over the attitudes and behaviors of those in their social networks.

Network analysis techniques can measure *innovation thresholds*, which define the number of people that need to sign on to something before the individual or community will adopt the change. Innovation thresholds can have significant implications for the design of a campaign. Another use for network analysis is to measure social capital and other constructs, like trust in the government or in adversary institutions.[9]

In terms of assessment, network analysis can inform the research process and sample selection strategy, including identifying reliable and valuable sources of information and input during the formative phase.[10] In the summative phase, network analysis can be used to track progress over time.

Target Audience Analysis

Effective audience analysis, known in the defense community as *target audience analysis* (TAA),[11] is the "cornerstone" of effective influence because it uncovers "root causes" and identifies the most effectual "levers to pull," in the words of one defense expert.[12] The basics of the process are laid out in doctrine; we briefly summarize the approach in the accompanying desk reference.[13] The information environment evolves rapidly. To effectively inform campaign planning, TAA should be conceived of as a living process rather than as a static picture of the information environment.

Further Reading

In this handbook:

Chapter Two, in the section "Evaluating Change Requires a Baseline," discusses the importance of baseline data for evaluating change.

Chapter Three, in the section "Three Types of Evaluation: Formative, Process, and Summative," addresses the role of formative assessment in identifying baselines.

Chapter Six, in Box 6.1, "Where to Begin? Measuring Baselines and Variables," addresses the importance of measuring baselines and characterizing the information environment.

---

[9] Author interview with Craig Hayden, June 21, 2013.

[10] Author interview with Simon Haselock, June 2013.

[11] Some communication experts, such as Thomas Valente, argue that DoD should consider moving away from the term *target* to describe an audience, because the term is perceived poorly by populations, particularly in a military context. On the other hand, incorporating audience analysis into the standard DoD targeting process would help integrate IIP activities with all military operations and processes.

[12] Author interview on a not-for-attribution basis, January 23, 2013.

[13] See, for example, Headquarters, U.S. Department of the Army, and Headquarters, U.S. Marine Corps, *Psychological Operations, Tactics, Techniques, and Procedures*, Field Manual 3-05.301/Marine Corps Reference Publication 3-40.6A, Washington, D.C., December 2003, chapt. 5. Alternatively, see Headquarters, U.S. Department of the Army, *Military Information Support Operations*, Field Manual 3-53, Washington, D.C., January 2013b.

In the accompanying desk reference:

Chapter Eight, in the section "Characterizing the Information Environment: Key Audiences and Program Needs," offers more detail on innovation thresholds, sociometric segmentation (social network analysis), and examples of the techniques addressed here. It also elaborates on the connection between TAA and content analysis/atmospherics.

Developing and Testing the Message

Developing the Message

After characterizing the IE, the next major task of formative research is to inform the development of the message or product. To develop effective messages, it is useful to solicit input from as many relevant sources as possible—for example, cultural anthropologists, ethnographers, trained participant observers, trusted local sources who understand the dynamics on the ground, and, if feasible, individuals from both sides in a conflict. Joshua Gryniewicz, the communication director at CureViolence, says his organization relies on neutral groups when adapting its model to local conditions. Neutral groups are not affiliated with a particular militia group or sect and are perceived as credible by all sides in a conflict.[14]

Testing the Message

Rigorously pretesting messages on a representative sample of the intended audience will dramatically improve the likely effectiveness of the message and will mitigate the chance of failure or unintended consequences. For example, a message designed to make tobacco use look "uncool" to teens could easily backfire if they perceive manipulation by adults. Likewise, DoD information or influence messaging must walk a fine line between promoting U.S. interests and being perceived as culturally insensitive. Testing the message in the formative phase is the best way to calibrate the messaging such that it achieves an effect without offending the audience. Piloting the intervention on a small scale can help refine the logic model, preemptively identify sources of program failure, and allow practitioners to fine-tune the message or the campaign. Despite the rich information provided by pilot programs, planners must keep in mind the different conditions for success at different scales. For example, will a message tested only regionally succeed in reaching key audiences at a national level?

Another way to test a message is in a "laboratory" setting. Psychological models of influence are often used to design the campaign, but the models are rarely validated or tested against results observed in the field.

Further Reading

In the accompanying desk reference:

Chapter Seven discusses split, or A/B, testing, which involves employing two variants of a message to two groups within the same audience segment and measuring differences in responses. This can be an effective message-testing technique in the formative phase.

---

[14] Author interview with Joshua Gryniewicz, August 23, 2013.

The Importance and Role of Qualitative Research Methods

Given the inevitable challenges associated with collecting valid and reliable quantitative data on IIP effects, evaluators should consider the balance between qualitative and quantitative information at all stages of evaluation. The best quantitative methods are those that supplement the information produced from qualitative methods, and vice versa.

Military analysts often prefer quantitative data not because such data are inherently more objective but because they are easier to analyze and they provide, in Jonathan Schroden's words, a "façade of rigor."[15] However, *numeric data* are not the same as *objective data*. Quantitative data are only as valid and reliable as the instruments and processes that generated them. Moreover, quantitative data are often *less* useful than qualitative data because they encourage data customers to view results as countable phenomena, which, in an IIP setting, are more likely to be associated with outputs than with meaningful outcomes.[16] Qualitative methods also help interpret or explain quantitative data, especially unexpected or surprising results. Qualitative methods are also better for determining causality and uncovering motivations or the drivers of change.[17]

Of course, qualitative data should be generated by rigorous social science methods. As one expert joked, "The plural of *anecdote* is not data."[18] Moreover, while qualitative methods add value to quantitative approaches, programmers should avoid making decisions on the basis of a single qualitative method.[19] Here, we briefly profile the advantages and challenges of a handful of the most common qualitative research methods.

Focus Groups

Focus groups are particularly valuable for testing products and anticipating how the audience will react to various dimensions of a product—message, imagery, language, music, and so forth. Matthew Warshaw recalls a few cases in which planned IO programs were canceled because focus groups showed that the message was "culturally insensitive or that the psychological objective [he was] seeking was flawed."[20]

There are several challenges to implementing focus groups in operational environments. First, they can be difficult to organize and require skilled local facilitators who share demographic characteristics with the focus group sample. Second, responses can be biased due to groupthink and normative pressures of conformity. In Afghanistan,

---

[15] Jonathan Schroden, "Why Operations Assessments Fail: It's Not Just the Metrics," *Naval War College Review*, Vol. 64, No. 4, Fall 2011, p. 99.

[16] Author interview with Simon Haselock, June 2013.

[17] Author interview with Matthew Warshaw, February 25, 2013.

[18] Author interview on a not-for-attribution basis, December 15, 2013.

[19] Author interview with Kim Andrew Elliot, February 25, 2013.

[20] Author interview with Matthew Warshaw, February 25, 2013.

Warshaw found that people tended to agree with each other and would encourage the group to come to consensus. Finally, outcomes can be unpredictable, and results are difficult to standardize and analyze.[21]

To manage these challenges, it is important to employ best practices for conducting focus groups, drawn from social sciences research. The accompanying desk reference offers a full list of these techniques.

Interviews

Like focus groups, one-on-one interviews can be used to test products, identify causal mechanisms, explain program failure, and validate and interpret survey results. Some researchers believe that these interviews are even better than focus groups for understanding causal mechanisms in conflict environments, because they avoid the challenges associated with groupthink and pressures to conform to social norms. Rapport between the interviewer and the respondent is very important. Interviewers should share characteristics with the subject and should begin the interview with noncontroversial subjects.[22]

Qualitative interview methods include in-depth interviews and intercept interviews. In-depth interviews are semistructured interviews between researchers and members of the target audience. Intercept interviews, or person-on-the-street interviews, are solicited in public places, such as a bazaar, and are useful for gauging public perceptions about a product or an issue. To get the most out of intercept interviews, researchers should pretest the instrument and vary the days, times, and interviewers.[23] While it is difficult to impose a formal sampling strategy, the sample of respondents should be as random as possible.

Narrative Inquiry

Narrative inquiry, or narrative analysis, is an approach for determining how members of a target audience create meaning in their lives through storytelling; it is not a primary method of data collection. It typically involves coding qualitative data collected through content analysis and qualitative methods (e.g., interviews and focus groups) using a standardized index. Cognitive Edge, Inc., has developed the SenseMaker software package that claims to be able to identify which attitudes have the potential to be changed and which do not. The tool processes a large volume of micronarratives collected from volunteer subjects, and then interprets, categorizes, and tags the stories into abstract categories.[24] While this method produces less valid and generalizable results

---

[21]   Author interview with Thomas Valente, June 18, 2013.

[22]   Valente, 2002, p. 58.

[23]   Valente, 2002 p. 60.

[24]   To read more about SenseMaker software, see SenseMaker, homepage, undated. Also see NATO Joint Analysis and Lessons Learned Centre, 2013, p. 42.

than a large, formal survey, it is less expensive and quicker, capable of providing real-time content directly from the target audience.[25]

On the analysis side, narrative is one way to make sense of disparate data, and to aggregate across programs, activities, and analyses of different types is to tell a compelling story. This method of analysis and aggregation is referred to as a *narrative approach* and has been strongly advocated for aggregate campaign and operational-level assessments by our RAND colleague Ben Connable.[26] Compiling information in a narrative can be viewed as a sort of holistic triangulation, interpreting all available data and making a compelling argument for its interpretation.

If a narrative analysis is conducted within the context of an explicit theory of change, it can contribute to assessment in important ways. For a narrative to have such a connection, it need not ever say "theory of change," but it must make a clear statement about how the various operations and activities being analyzed are supposed to connect to desired end states, describe progress toward those end states, and offer an explanation of any shortfalls in progress.

However, like all assessments, where underlying data are suspect, resulting narratives can be suspect. Of course, if the analyst or narrator is aware of weaknesses in the underlying data, that can become part of the narrative and thus an analytic strength. And like self-assessment of any kind, narratives are vulnerable to bias and overoptimism. Although narratives can pose challenges, their advantage is in allowing analysts to capture variations and nuances across the area of operations; they can also remind stakeholders of the context and complexity of an operation, force assessors to think through issues and ensure that their assessment is based on rigorous thought, and ensure a proper balance between quantitative and qualitative information, between analysis and judgment, and between empirical and anecdotal evidence.[27] See the additional discussion in Chapter Eleven of narrative as a means of presenting assessment results.

Anecdotes

Anecdotes are widely used to communicate the effectiveness of IIP programs. Sometimes, anecdotes are used because a more rigorous measurement system is not in place. In other cases, measures are not perceived as necessary because the effect is supposedly evident. Anecdotes are not just easier to generate than experimental evidence; they are often more powerful.

But anecdotes are often used to demonstrate effect even when more-rigorous measures *are* available. Anecdotes alone are insufficient to empirically demonstrate impact because there is no counterfactual condition to infer causality and no basis on which to

---

[25] NATO, Joint Analysis and Lessons Learned Centre, 2013, p. 42.

[26] Ben Connable, *Embracing the Fog of War: Assessment and Metrics in Counterinsurgency*, Santa Monica, Calif.: RAND Corporation, MG-1086-DOD, 2012.

[27] Schroden, 2011, p. 99.

generalize. However, it is good practice to embed stories or narratives into the presentation of the evaluation results to give meaning or color to the quantitative measures.[28]

Expert Elicitation

While eliciting expert judgment is considered methodologically inferior to experimental designs, in many circumstances, structured expert elicitation is the most rigorous method among all feasible and cost-effective options. Eliciting expert judgment can take many forms, from informal BOGSATs to highly structured, iterative Delphi processes requiring consensus and insulation from personality or authority.[29] The accompanying desk reference discusses two expert elicitation methods used to inform IIP assessment: the Delphi method and interviews with commanders.

Other Methods

In our interviews, we heard about three other qualitative techniques commonly used in the private sector: community assessments, temperature maps, and participatory photojournalism. Community assessments target disadvantaged or vulnerable populations and encourage them to express issues visually or in their own words. Temperature maps are visual representations of issue saliency across geographic areas. In participatory photojournalism, subjects are asked to take pictures of the things that matter to them, and the results are used to gauge perceptions of governance.[30]

SMEs also discussed the cultural consensus method, which measures shared knowledge or opinions within groups. It is used in conjunction with focus groups and in-depth interviews to uncover the core of an issue while attempting to gain an understanding of the atmospherics and perceptions in different provinces.[31]

Further Reading

In the accompanying desk reference:

Chapter Eight offers more detail on best practices, including examples from across the sectors considered in this research, for each of the qualitative research methods described here.

Chapter Nine, in the section "Narrative as a Method for Analysis or Aggregation," elaborates on the role of narrative in analysis and data aggregation.

Chapter Eleven explains the role of narrative in presenting and facilitating the understanding of aggregated data in assessment results.

---

[28] Author interview on a not-for-attribution basis, July 31, 2013.

[29] *BOGSAT* is a nonstandard but common acronym for "bunch of guys sitting around a table," not a particularly rigorous approach to expert elicitation.

[30] Author interview with Kavita Abraham Dowsing, May 23, 2013.

[31] Author interview on a not-for-attribution basis, March 2013.

Key Takeaways

- To construct effective messages, planners must understand what messages and what formats resonate with what audiences. Audiences should be segmented according to psychographic variables and their demographic correlates rather than strictly by demographics.
- In some cases, campaigns may use an indirect-effects strategy that targets influencers of the focal audience (e.g., family members, religious leaders). Social network analysis should be used to identify key influencers and opinion leaders.
- TAA should be understood as a living process rather than a static picture and should use up-to-date data on target audience sentiments to shape messages right up to the point of dissemination.
- Messages and products should be pretested with qualitative techniques (e.g., focus groups) or with more-rigorous, more-controlled methods (laboratory experiments).
- Piloting the intervention on a small scale and using computer-generated simulations can help refine the logic model and preemptively identify sources of program failure.
- The plural of *anecdote* is not *data*. Qualitative data should be generated by rigorous social science methods. Likewise, decisionmakers should not be expected to make decisions on the basis of a single quantitative method.

Surveys and Sampling in DoD IIP Assessment
Best Practices and Challenges

Survey research is a useful and efficient method for gathering information on the traits, attributes, opinions, and behaviors of people.[1] Survey research can serve as a valuable tool for IIP efforts by providing needed information regarding a population of interest or permitting measurement of the effects (or lack of effect) of an implemented program. However, surveys are not without limitations, and various sources of error can hinder the collection of reasonably accurate information. For example, error can arise from badly designed survey items, poorly translated surveys, and surveys that have been administered incorrectly.[2] Another source of error can be the collection of survey data from a particular sample, or a portion of the population, that does not adequately represent the whole population of interest.

Best Practices for Survey Management

Before addressing sample selection, survey instrument design and testing, and the uses of survey data, we briefly discuss the management and oversight of survey research in support of IIP activities. Survey programs are complex, with many moving parts. Successful implementation requires vigilant oversight across the entire process, input from experts and stakeholders, and a willingness to collaborate and be scrutinized.

Those responsible for contracting, staffing, or overseeing the administration of a survey in support of IIP assessment should consider the following recommendations.

- Engage and involve cultural experts, survey research experts, stakeholders, and other organizations familiar with the target audience. These experts can help with vetting local research firms, designing and testing the survey

---

[1] Don A. Dillman, Jolene D. Smyth, and Leah Melani Christian, *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*, 3rd ed., Hoboken, N.J.: John Wiley and Sons, 2009.

[2] Maureen Taylor, "Methods of Evaluating Media Interventions in Conflict Countries," paper prepared for the workshop "Evaluating Media's Impact in Conflict Countries," Caux, Switzerland, December 13–17, 2010.

instrument, selecting the sample, and charting the logistics of the survey admin-
istration.

- Involve locals in the design of the survey instrument.
- Maintain continuity in survey management. It is better to have reachback man-
  agement with deployed operational analysts rather than charging deployed per-
  sonnel with the task.[3]
- Ensure that data collectors represent the demographics of the respondents.
  Depending on the environment, survey personnel may need to be matched
  according to religion, age, and local dialect.[4]
- Thoroughly vet local research firms prior to awarding contracts. Pressure to give
  contracts to the lowest bidder can lead to quality-control challenges.
- Keep records of high- and low-performing research firms to ensure that low-
  performing firms or firms caught cheating are not rehired when a contracting
  officer rotates in.
- Make an up-front investment in building local research capacity. DoD IIP cam-
  paigns will benefit in the long run by saving the costs associated with redoing
  surveys.[5]
- The initial contract with a survey research firm should cover one wave of poll-
  ing and be flexible. The contract should permit changes to the survey design and
  should include early termination clauses to prevent and manage cheating.[6]
- If the first survey is successful, subsequent contracts should seek to establish con-
  tinuity in survey design and a long-term relationship between the contracting
  unit and local research firm.

There is a widely perceived lack of transparency and "aversion to cooperation and
sharing" that creates inefficiencies and duplication in survey research in environments
like Afghanistan.[7] To avoid "reinventing the wheel," share survey data and results, and
leverage work done by others, whenever possible.[8]

Further Reading

In the accompanying desk reference:

Chapter Four, in the section "Cultivating Local Research Capacity," discusses the importance of building
local research capacity, including examples of where this has been done successfully.

Appendix B, in the section "Survey Management, Oversight, Collaboration, and Transparency,"
addresses building local research capacity for surveys. That section also includes a discussion of
managing cheating by local firms.

---

[3]   Eles et al., 2012, p. 31.

[4]   Author interview with Amelia Arsenault, February 14, 2013.

[5]   Author interview with Amelia Arsenault, February 14, 2013.

[6]   Eles et al., 2012, p. 31.

[7]   Author interview with Kim Andrew Elliot, February 25, 2013.

[8]   Author interview with Amelia Arsenault, February 14, 2013.

Sample Selection: Determining Whom to Survey

One important goal of a great deal of survey research is to collect data that provide accurate estimates about a population. In other words, researchers would like their survey assessments to correctly capture the characteristics of the populations they survey. This section provides practical information regarding survey sampling that may help IIP planners obtain representative information from a population of interest.[9]

Collecting Information from Everyone or from a Sample

A *census* involves collecting data from all the people in the population of interest. However, most research in the social sciences involves collection of data from a sample of the population, rather than from every person in the entire population.[10] Results that approximate those that would have been obtained had data been collected from an entire population can be obtained from a small selection of people from the population, given a reasonable amount of statistical error. Thus, a large amount of money and time can be saved by collecting data from a well-considered sample, rather than by collecting a census.

Sample Size: How Many People to Survey

As noted, some error exists in terms of the extent to which a sample represents the population. In other words, the precision of a sample can vary. All else being equal, a larger sample means less error. Variability also drives sample size. For example, if individuals in a population hold very different opinions on a topic, a larger sample size will be needed to better capture the entire population's opinion on the topic. IIP planners should consider how much error they are willing to accept in terms of their survey estimates.

Another element to consider when determining from how many people to collect survey data is subsequent data analysis. Researchers want to be able to observe a relationship between variables. In other words, if there is an association to observe (sometimes there is not), they need enough statistical power to be able to observe that association and thereby find statistical significance. Usually, researchers want to have an 80-percent chance of detecting an effect if it is present.

Some individuals have provided rules of thumb regarding sample sizes for different assessment approaches (see Table 9.1).[11] These recommended sample sizes can be inaccurate, so researchers have created tools that allow others to more accurately deter-

---

[9] Arturo Muñoz, *U.S. Military Information Operations in Afghanistan: Effectiveness of Psychological Operations 2001–2010*, Santa Monica, Calif.: RAND Corporation, MG-1060-MCIA, 2012.

[10] William D. Crano and Marilynn B. Brewer, *Principles and Methods of Social Research*, 2nd ed., Mahwah, N.J.: Lawrence Erlbaum Associates, 2002.

[11] Mertens and Wilson, 2012.

Table 9.1
Approximate Sample Sizes as Based on Approach

Approach	Rough Approximation of Minimum Sample Size Required
Correlational	82 participants (two tailed)
Multiple regression	At least 15 participants per variable
Survey research	100 participants for each major subgroup: 20–50 for minor subgroups
Causal comparative	64 participants (two tailed)
Experimental or quasi-experimental	21 individuals per group (one tailed)

SOURCE: Adapted from Mertens and Wilson, 2012.

mine the number of people from which they should collect data. A popular and free tool that may be used is called G*Power.[12]

Challenges to Survey Sampling

There are many challenges to survey sampling, and they are often magnified in an operational setting. Here, we review two common problems: nonresponse and lack of access.

Nonresponse

Rarely do all those who are asked to complete a survey agree to participate. This can lead to differences between the group that was sampled and the group that actually responded, which can keep results from being representative even if the sample was selected in a representative way. For example, those who choose not to participate may have more-favorable attitudes toward the government, may be more likely to be male, or may be better educated. Thus, their responses may not represent the total population of interest. This is called *nonresponse bias*. In a conflict environment, nonresponse is especially problematic, as many potential participants may be concerned about the repercussions of their responses or even participating in a survey.

In determining the extent of nonresponse bias, researchers often calculate and report the response rate, which is the number of completed surveys divided by the total number of people asked to participate in a survey. Different strategies may be implemented to promote responses. For example, female survey administrators may assist in promoting response rates among females, and the provision of small incentives may

---

[12] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang, "Statistical Power Analyses Using G*Power 3.1: Tests for Correlation and Regression Analyses," *Behavior Research Methods*, Vol. 41, No. 4, November 2009. Also see Department of General Psychology and Occupational Psychology, Heinrich-Heine-Universität Düsseldorf, "G*Power: Statistical Power Analyses for Windows and Mac," web page, undated.

also increase response rate.[13] Keeping surveys at a reasonable length and guaranteeing the anonymity of participant responses have also been suggested.[14]

There are also several methods to reduce the impact of nonresponse bias after a survey is completed.[15] These often involve comparing information about respondents (e.g., location, gender) with known information about nonrespondents to see whether nonresponse appears to be systematic (and concerning) or random (and thus less so).[16]

Lack of Access

In conflict environments, surveys often must be administered in person, and lack of access is particularly problematic. For example, survey takers may be turned away, areas may be too difficult to reach, or areas may be too dangerous to enter.[17] Yet these areas are often those of greatest interest in IIP efforts. It is important to keep records on inaccessible areas so that they can be tried again or so that missing data can be accounted for when reporting results. It may also be necessary to realign the sampling frame as based on areas that are accessible and inaccessible.[18]

Interview Surveys: Surveying Individuals in a Conflict Environment

In-person interviews and phone interviews involve interviewers verbally asking each question, providing the response options for each question, and then recording the selected response. For a variety of reasons, this may be the only option available to survey planners in an operational area (see Box 8.1 in Chapter Eight for more details).[19]

Interview surveys can be costly and timely because interviewers must sit with each person,[20] but interview surveys have several advantages over self-administered surveys. They often have higher response rates than self-administered mail surveys, especially in conflict environments,[21] and they may produce more reliable and less biased results.[22]

---

[13] Author interview with Matthew Warshaw, February 25, 2013.

[14] Crano and Brewer, 2002.

[15] J. M. Brick and G. Kalton, "Handling Missing Data in Survey Research," *Statistical Methods in Medical Research*, Vol. 5, No. 3, September 1996.

[16] Brick and Kalton, 1996; Joseph L. Schaefer and John W. Graham, "Missing Data: Our View of the State of the Art," *Psychological Methods*, Vol. 7, No. 2, June 2002.

[17] Author interview with Matthew Warshaw, February 25, 2013.

[18] Eles et al., 2012.

[19] In some environments, methods other than in-person approaches may be possible. The ubiquity of mobile phones in some countries has opened more opportunities for administering telephone surveys, and some groups have begun to use short message service (SMS) or text messages to administer surveys.

[20] Author interview with Emmanuel De Dinechin, May 16, 2013.

[21] Author interview with Matthew Warshaw, February 25, 2013.

[22] Author interview with Kim Andrew Elliot, February 25, 2013.

Administering surveys in-person may decrease the number of questions that respondents answer using the "don't know" or "refuse to answer" options, and interviewers can assist in addressing respondents' misunderstandings regarding survey items (but this must be strictly controlled). Finally, interviewers can record observations regarding the respondents and their surroundings, such as the characteristics of the dwelling and reactions of participants to certain survey items.[23]

However, different elements of survey interviews must be carefully considered. In survey interviews, the interviewer's presence and presentation of items should not influence, or should influence as minimally as possible, how each respondent interprets and then answers each survey item. The interviewer's tone, nonverbal cues, and characteristics are all elements that may influence participant responses. To address the influence of interviewer characteristics, some have suggested attempting to match the characteristics of the interviewer and respondent.[24] This may include matching race/ethnicity, first language spoken, religion, and gender of the interviewer and respondent.[25]

Box 9.1
Challenges to Sampling in a Conflict Environment

In addition to deciding whom to include in a focus group, survey, or set of interviews, IIP planners must also consider how they are going to collect data from these individuals. Data collection methods vary in terms of cost and information quality, and the method used should be appropriate for the population of interest.

In a conflict environment, it can be particularly difficult to obtain accurate contact information for targeted populations: People might move to avoid violence, they might be reluctant to register with authorities, they might not have access to reliable telephone or Internet service, or literacy levels may be low.

Other factors that can complicate sampling include the lack of a credible census, limited access to people in geographically challenging or dangerous areas, and an inability to speak with certain individuals, such as women or those who are not the head of a household.[a] These and other data collection constraints can lead to unrepresentative samples and other types of sampling errors.

A best practice in survey management in an operational context is to match the sample with interviewers or survey takers who are demographically alike. This can prove challenging in that it is often difficult to find willing individuals who have the required characteristics and are literate. A related challenge encountered by U.S. government programs has been quality control when employing local firms. Faulty record keeping and other uncertainties in conflict areas can make it difficult to vet firms, and it is not unusual for firms with poor track records to repeatedly compete for and even win new contracts.

Despite the potential difficulties in addressing sources of error in a conflict environment, surveys continue to be used, in part, because they provide information that can be presented to and used by military commanders and Congress.

[a] Author interview with Kim Andrew Elliot, February 25, 2013.

---

[23] Earl Babbie, *Survey Research Methods*, 2nd ed., Belmont, Calif.: Wadsworth Publishing Company, 1990.

[24] Babbie, 1990.

[25] Author interview with Amelia Arsenault, February 14, 2013.

In addition, the survey interviewers should be well trained in how to administer a survey. There are various rules for survey interviewing, stipulating, for example, that an interviewer's appearance and demeanor should somewhat correspond to those being interviewed (e.g., an interviewer should dress modestly when interviewing poorer respondents).[26] Further, interviewers should be very familiar with the questionnaire so that they can read items without error. They should also read questions exactly as written and record responses exactly as provided. When surveys are administered in the field, there should be a clear plan for supervisor oversight.

Further Reading

In the accompanying desk reference:

Chapter Ten discusses survey sample selection in greater detail.

The Survey Instrument: Design and Construction

Here, we review some best practices in survey design. Later in this chapter, we discuss how to mitigate bias when administering a survey. When IIP assessment planners design (or contract for) surveys, they must consider question wording and overall survey length, question structure, question order, and response options.

Question Wording and Survey Length: Keep It Simple

Questions that are simpler are more likely to be understood by respondents.[27] Complex or vague questions that attempt to indirectly assess a certain topic can contribute to respondent confusion and reduce the utility of responses.[28] As such, questions should be short and use simple terms.[29]

Surveys should always avoid double-barreled questions, in which respondents are asked about two concepts in one question and are allowed to provide only one response. For example, the question "Do you think certain groups have gone too far and the government should crack down on militants?" addresses two concepts: the behavior of certain groups and the desired behavior of the government. A response to this question may be addressing either of these two concepts, but which one cannot be determined. This uncertainty makes it difficult to code the survey results.

In addition to asking simple questions, it is important to keep the survey as short as possible. Survey fatigue occurs when respondents lose interest, and their motivation to complete a survey wanes. This can occur when a survey is too long or complex, or when the same person has been asked to participate in multiple surveys. One way to

---

[26] Babbie, 1990.

[27] Crano and Brewer, 2002.

[28] Taylor, 2010, p. 10.

[29] Valente, 2002, p. 124.

prevent survey fatigue is to inform participants how long it will take to complete a survey; they may be less likely to experience fatigue when their expectations have been set prior to starting the survey.[30] To avoid burdening the same people with similar surveys, the other cause of survey fatigue, it is useful to determine whether similar data are already being collected by other organizations and asking to share data with those groups.[31]

Open-Ended Questions: Added Sensitivity Comes at a Cost

Open-ended questions involve asking respondents a question and then allowing them to provide their own answers. For example, an open-ended question might ask, "Who is your favorite presidential candidate?" A closed-ended version of this question would be worded the same way but would provide a limited set of response options. Asking open-ended questions can capture information that would not otherwise have been. The format also allows respondents to explain their responses.[32]

However, open-ended questions come with costs. It takes respondents longer to provide responses to open-ended questions. This increases the participant's time commitment and may increase the likelihood of survey fatigue.[33] In addition, it can be difficult to capture participants' responses accurately, and interpreting and analyzing open-ended responses can be a complex and onerous process that requires the creation of a reliable coding scheme.[34] These questions should be used sparingly, when questions have no clear set of predefined answer options or when more-detailed responses are needed.

Question Order: Consider Which Questions to Ask Before Others

When implementing a survey, respondents who feel comfortable with and committed to the research may be more likely to respond to sensitive questions.[35] To establish comfort and build rapport, the least-threatening survey items should be asked at the beginning of the survey. Once respondents have answered these, they may be more willing to respond to later questions that may be perceived as more personal or threatening. Do not assume that demographic questions are least threatening, however. Income, education level, and marital status may all be sensitive topics, and these questions may raise privacy concerns for respondents. Instead, easy-to-answer questions that are relevant to the survey may be best to present first.

---

[30] Dillman, Smyth, and Christian, 2009.

[31] Eles et al., 2012.

[32] Author interview on a not-for-attribution basis, March 1, 2013.

[33] Dillman, Smyth, and Christian, 2009.

[34] Eles et al., 2012.

[35] Crano and Brewer, 2002.

In addition, a person's responses to earlier questions can influence his or her responses to later questions. For example, if a number of questions ask respondents about the influence of terrorism on their country and a subsequent open-ended question asks what they believe to be one of the biggest threats to their country, terrorism may be a more likely response than it would have been had the open-ended question been asked first.[36] To control for this influence, the research recommends creating more than one version of a survey, varying the order of items or sets of items.[37] When using this technique, the least-threatening items should remain at the beginning of the survey.

Survey Translation and Interpretation: Capture the Correct Meaning and Intent

Surveys developed for U.S. government efforts are often written in English and then translated into the local language before being fielded. Without proper review, the original meaning and intent may be lost in translation.[38] Back-translation is one way to correct for translation errors. In back-translation, a translated survey is translated *back* into its original language (by someone other than the original translator).[39] The back-translated survey should match the original as closely as possible. Back-translation can reveal, for example, words that are literally equivalent in two different languages but may not have equivalent meanings.[40]

One thing back-translation might not do, however, is indicate whether certain groups may take offense to the wording of certain items, such as items regarding women's rights and perceptions of elders.[41] To reduce this possibility, surveys should be reviewed by individuals who are local to the area to be surveyed.[42]

Multi-Item Measures: Improve Robustness

Surveys often seek to address complex concepts, and a single survey item may not adequately address a complex concept. For example, to assess religiosity, a survey may include an item asking about frequency of mosque or church attendance. However, those who frequently attend mosque or church may not appear as strongly religious if their answers on subsequent questions about frequency of prayer or strength of certain beliefs show that, say, they do not pray very often or they do not embrace certain tenets

---

[36] Babbie, 1990.

[37] Crano and Brewer, 2002.

[38] Eles et al., 2012.

[39] Robert Rosenthal and Ralph L. Rosnow, *Essentials of Behavioral Research*, 3rd ed., New York: McGraw-Hill, 2008.

[40] Martin Bulmer, "Introduction: The Problem of Exporting Social Survey Research," *American Behavioral Scientist*, Vol. 42, No. 2, October 1998.

[41] Eles et al., 2012.

[42] Author interview with Amelia Arsenault, February 14, 2013. Surveys can be vetted through the use of focus groups and other techniques.

of their religion.[43] As such, it is often worthwhile to utilize more than one item to assess a construct. Collectively, these items are called an *index* or *scale*.[44] If all of the items in a scale assess the same construct, these items can be aggregated. Scales can provide more-comprehensive and reliable measures of complex concepts than use of single items.[45]

There are several types of scales, but one of the most common is a Likert scale.[46] With this method, participants are presented with several items on a topic and can choose one of several responses to each item, presented as a range. For example, a survey might ask participants the extent to which they agree with the following statement: "The national government has had a positive influence on my life." Participants could then indicate their level of agreement using one of five possible response options (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, and 5 = strongly agree). Several additional items addressing perceptions of the national government may be asked and, then these items may be summed or averaged together. Before combining responses to items, it is important to determine the extent to which the items are related. If items are positively related, that suggests they are measuring the same construct. One way to assess whether scale items are sufficiently related is by calculating an alpha coefficient. When using scales, it is important to keep in mind the risk of survey fatigue. Ask only as many questions as necessary to obtain the information you require.

Item Reversal and Scale Direction: Avoid Confusion

The simplest surveys consist of items with parallel constructions. That is, questions are posed in a similar way and the response options are the same across all questions. Sometimes, survey developers opt to include questions that follow a different format, solicit a different type of response, or request that respondents relay their responses using a scale that moves in the opposite direction. This is often done for lack of a better approach to collect the information required, but asking the exact question you need to ask to obtain the exact information you require has a downside: Changing formats and scales may confuse participants, increasing the risk that you will get inaccurate data anyway. Further, items that need to be reversed before being combined with other items in indexes or scales risk being reversed more than once between collection and final analysis. This leads to two suggestions: (1) where possible, avoid reverse-scale items, and (2) always protect and preserve the raw data so that any analytically driven recoding can be tracked and undone, if necessary.

---

[43] Babbie, 1990.

[44] Valente, 2002, p. 151.

[45] Author interview with Ronald Rice, May 9, 2013.

[46] Ronald J. Thornton, "Likert Scales: An Assessment Application," *IO Sphere*, Summer 2013.

Further Reading

In the accompanying desk reference:

Chapter Nine, in the section "Embedding Behavioral Measures in Survey Instruments," addresses the use of surveys to measure how people actually behave (revealed preferences) in addition their stated preferences.

Response Bias: Challenges to Survey Design and How to Address Them

A number of factors may influence participant responses to survey items, including interviewer characteristics and question ordering. Ideally, researchers would like the characteristics of the survey to have a minimal influence on responses. However, this can be difficult, and survey designers should be aware of factors that influence participant responses.

Social Desirability Bias

One potential threat to capturing respondents' true attitudes and perceptions is known as *social desirability bias*—when people try to present themselves in a manner that their society regards as positive.[47] Rather than responding to an item or set of items in a way that reflects their true perceptions or actual attitudes, participants may instead respond based on how they believe that their society would like them to respond. This distorts participant responses and researchers' ability to better understand attitudes and perceptions.

   To address this, some suggest inclusion of a ten-item social desirability scale in the administered survey. Responses to certain survey items that are strongly correlated with participants' responses on the scale may suggest survey items that should be excluded from analyses.[48] Informing participants that their responses are anonymous may also increase candor, reducing the influence of social desirability bias.[49]

Response Acquiescence

Another factor that may distort participant responses is known as *response acquiescence*. Other terms for this same concept include *agreement bias* or *response affirmation*. Response acquiescence occurs when survey respondents agree with survey items, regardless of the content.[50] Thus, given a set of survey items and asked to respond on a scale ranging from 1 (strongly disagree) to 5 (strongly agree), respondents will tend to express higher levels of agreement without thoroughly considering what they are agreeing to.

---

[47] Robert F. DeVellis, *Scale Development: Theory and Applications*, 3rd ed., Thousand Oaks, Calif.: Sage Publications, 2012.

[48] DeVellis, 2012.

[49] Crano and Brewer, 2002.

[50] Crano and Brewer, 2002.

To address this, researchers include both positively and negatively worded items within a scale. For example, if assessing self-esteem, researchers may include items focused on high self-esteem (e.g., "I feel that I have a number of good qualities") and items focused on low self-esteem (e.g., "I feel useless at times").[51] The responses of someone who tends to agree with all items, regardless of content, would be balanced across survey items, revealing their response acquiescence. Unfortunately, using positively and negatively worded items may confuse respondents and analysts. (See the section "Item Reversal and Scale Direction," earlier in this chapter.)

Mood and Season

An additional factor that may influence participant responses is their mood, which may be associated with the season. For example, previous research has shown that participants respond more negatively when it is raining than when it is sunny.[52] Other researchers have noted that participants in conflict environments may have difficulty finding fuel for cooking or keeping warm in the winter, which may dampen their general outlook.[53]

To address the influence of season and mood on responses, researchers should consider collecting data at different times of the year and assessing patterns in responses across these periods. Another strategy is to first ask participants questions about the weather, which may decrease the likelihood that they will incorrectly attribute their negative feelings to their general life situations rather than the bad weather.[54] (Questions about the weather also have the added bonus of being nonthreatening and thus ideal for inclusion at the beginning of a survey; see the section "Question Order," earlier in this chapter.)

Further Reading

In the accompanying desk reference:

Chapter Ten discusses each type of response bias in greater detail in the section "Response Biases: Challenges to Survey Design and How to Address Them."

Testing the Survey Design: Best Practices in Survey Implementation

This chapter has so far focused on actions that IIP planners can take to address specific challenges that can arise during survey design and implementation, but best practices favor the systematic assessment of the survey at every stage in the process, including

[51] DeVellis, 2012.

[52] Norbert Schwarz and Gerald L. Clore, "Mood, Misattribution, and Judgments of Well-Being: Information and Directive Functions of Affective States," *Journal of Personality and Social Psychology*, Vol. 45, No. 3, September 1983.

[53] Author interview with Matthew Warshaw, February 25, 2013.

[54] Crano and Brewer, 2002.

after the survey is administered. Many of the techniques for testing a survey design parallel those recommended for testing a campaign message.

Pretesting

Before implementing a full-scale survey, survey designers should determine whether the target sample will understand the questions, whether they are able to respond to the questions, and whether interviewers can appropriately administer the survey in relevant social context.[55] Different avenues available for pretesting a survey include focus group discussions and individual interviews in which participants respond to the survey and explain what they were thinking when responding to each item. After the survey is modified according to this feedback, pilot testing (administering a small number of surveys) in the field can begin.[56] Pretesting and pilot testing can help address potential issues before the costly, wide-scale implementation.

Maintaining Consistency

At times, commanders or IIP planners may seek to assess changes in attitudes or perceptions. To do so, it is typically necessary to administer surveys over a long period of time.[57] These surveys should use the same wording and the same response options so that changes in responses can be assessed over time. Changing the wording, response options, or scales hinders the assessment of changes in attitudes. This is another case in which rotations can cause challenges in the operational environment: If a new commander seeks to measure different constructs, these changes should be carefully considered, because consistency and continuity will permit better assessments of change.[58]

Review of Previous Survey Research in Context of Interest

When developing a new survey to be administered to a given population, IIP planners should review previous research that has been conducted in the area and previous research that has been conducted on the topics of interest. Multiple examples of survey research are available and may be consulted for this purpose. These include Altai Consulting's assessment of Afghan Media in 2010, YouGov data collected in Iraq, and various research efforts conducted by the British Council.[59]

---

[55] Floyd J. Fowler, *Improving Survey Questions: Design and Evaluation*, Thousand Oaks, Calif.: Sage Publications, 1995.

[56] Fowler, 1995.

[57] Eles et al., 2012.

[58] Eles et al., 2012.

[59] See Altai Consulting, "Afghan Media in 2010," prepared for the U.S. Agency for International Development, 2010. The synthesis report and supplemental materials, including data sets and survey questionnaires, are available on Altai Consulting's website. See also UK Polling Report, "Support for the Iraq War," online database, undated, and British Council, *Trust Pays: How International Cultural Relationships Build Trust in the UK and Underpin the Success of the UK Economy*, Edinburgh, UK, 2012.

Using Survey Data to Inform Assessment

After survey data have been collected, they must be analyzed, triangulated with other data sources, and interpreted so as to meaningfully inform IIP assessment. This section addresses these processes.

Analyzing Survey Data for IIP Assessment

This section offers broad, high-level recommendations for the analysis of survey data in support of IIP assessment in conflict areas. It does not address statistical procedures in detail.

To allow analysis of trends over time, all waves of the survey should be combined into a master data set. A failure to do so has complicated efforts to analyze polls in Afghanistan.[60] Statistical software, such as SAS, STATA, and R, can be used to merge multiple waves of survey data. Polling programs should use advanced statistical packages but should keep versions of the data sets in standard formats to facilitate sharing and transparency.[61] It is worth emphasizing here that the quantity and quality of the data are far more important than the analytical technique or software program used. Even the most-sophisticated techniques cannot overcome bad data.

The sampling error, often expressed as the margin of error, represents the extent to which the survey values may deviate from the true population values. As discussed in the section "Sample Size," earlier in this chapter, survey error is inversely related to sample size. In Afghanistan, nationwide surveys have margins of error of plus or minus 3 percent, and district surveys have margins of error closer to 10 percent.[62] Because less is known about the population in operating environments like Afghanistan, survey research should continuously inform estimates of design effects and associated margins of error. When data from multiple surveys are available, analysts should examine variation across variables that should be constant (e.g., age, marital status) to revise estimated survey errors.[63]

Analyzing and Interpreting Trends over Time and Across Areas

Survey results can shape how decisionmakers perceive trends over time and across regions. The best surveys in support of IIP assessment are those conducted in several areas and repeated frequently over time. This is true for several reasons. First, surveys in conflict environments are particularly prone to response and nonresponse bias. Analyzing data over time and across areas controls for these sources of bias, assuming that

---

[60] Eles et al., 2012, p. 37.

[61] Eles et al., 2012, pp. 36–37.

[62] Downes-Martin, 2011, p. 110.

[63] Eles et al., 2012, p. 36.

they are not correlated with time or region.[64] Second, repeated measurement provides a means to validate the survey by assessing whether observed shifts in attitudes reflect expected relationships with known or likely triggers of attitudinal change, such as upticks in violence or kinetic operations, civilian casualties, or political turmoil.

Triangulating Survey Data with Other Methods to Validate and Explain Survey Results

Given the large margins of error and challenges posed by nonresponse and response bias, survey data are most valuable to IIP assessment when analyzed over time and in conjunction with other qualitative or quantitative data sources. Evaluators should validate survey results by assessing whether data or indicators produced by other methods are trending in the same direction or converging with survey data. This point was made by nearly every expert interviewed for this study with experience conducting or using surveys in conflict environments.[65]

In addition to validating survey results, other methods—particularly qualitative methods—should be used to explain and interrogate survey results, especially if they are unanticipated. It is often stated that the survey data tell you *what* and the qualitative data tell you *why*.[66] Thomas Valente characterizes the relationship between qualitative methods and survey research as an *iterative process*: Qualitative research informs the design of the survey, and the survey generates questions that are probed by a second iteration of qualitative research, which feeds into the revision of the survey instrument.[67]

Key Takeaways

- Those responsible for contracting, staffing, or overseeing the administration of a survey in support of IIP assessment should adhere to best practices for survey management, including engaging experts and local populations in survey design, vetting and tracking the performance of local firms, and maintaining continuity throughout the survey period.
- IIP planners should consider whom they would like to survey, how many people to survey, and what procedure to use to administer the survey. Survey takers should represent the target population as closely as possible.

---

[64] Eles et al., 2012, pp. 37–38.

[65] Author interview with Simon Haselock, June 2013; author interview with Jonathan Schroden, November 2013; author interview with Steve Booth-Butterfield, January 2013; author interview with Maureen Taylor, April 2013.

[66] Author interview with Maureen Taylor, April 2013.

[67] Author interview with Thomas Valente, June 18, 2013.

- When considering the ideal number of people from whom to collect survey data, IIP planners should keep in mind the variability in the attitudes and behaviors of the population of interest. Generally, greater variability warrants larger sample sizes.
- Nonresponse and lack of access are challenges inherent in all survey efforts. This is especially true for survey efforts conducted in conflict environments, where populations may move frequently, people may lack access to telephones or the Internet, and areas are inaccessible.
- Surveys should be designed such that the instrument or collection methods do not greatly influence participant responses. Question wording and overall survey length, question structure, question order, and response options can all affect participants' responses.
- Social desirability bias (a desire to conform to social expectations), response acquiescence (a tendency to agree with questions, regardless of their content), and even the respondent's mood, the season, or the weather can affect responses.
- Best practices in survey design and implementation favor the systematic assessment of the survey at every stage in the process, including after the survey is administered.
- Triangulating survey results, comparing a survey's results with information obtained from other surveys or focus groups, may also assist with survey validation.

Measurement
Collecting IIP Outputs, Outcomes, and Impacts

This chapter describes the methods that help decisionmakers answer one of the core questions motivating this report: Is an IIP effort working? We begin with an overview of research methods and discuss the importance of data quality and quantity. We then describe the methods and data sources for process evaluation.

Overview of Research Methods for Evaluating Influence Effects

The primary research methods and data sources for evaluating IIP effects are surveys; content analysis, including traditional media monitoring, web analytics, and social media monitoring and frame analysis; direct observation, or atmospherics; network analysis; direct response tracking; and qualitative methods, including focus groups, in-depth interviews, narrative inquiry, and Delphi panels. Secondary and aggregate data, such as data on economic growth or casualties, can also inform summative evaluations. Anecdotes and self-assessment, in which commanders evaluate progress made by subordinate units, are commonly used informal methods for gauging effectiveness.

NATO's framework for assessing public diplomacy summarizes several of these methods in a table that maps each method to resources required and a time frame for results. A modified version of this menu of research methods is presented in Table 10.1.

Further Reading
In this handbook:

Chapter Eight discusses formative and qualitative research methods in a general sense.
Chapter Nine presents best practices for survey development to facilitate the process of populating assessments with survey results.

In the accompanying desk reference:

Chapter Nine presents a more in-depth overview of research methods and data sources for evaluating DoD IIP efforts, including secondary and aggregate data sources.

Table 10.1
Menu of Research Methods for Assessing Influence Activities

Research Method	Role in Preintervention Evaluation	Role in Postintervention Evaluation	Resources Required	Validity	Time Frame for Results	Manpower Requirements	Limitations
Representative survey	Characterize IE and baseline	Measure exposure and attitudes	High	High	Immediate to several weeks	Survey research group, locals	Access, nonresponse, and response bias
Content/sentiment analysis: traditional media	Characterize IE	Measure distribution and changes in attitudes and beliefs	Medium	Medium high	Weeks	Outsource, local coders	Unrepresentative samples, difficult to code
Content/sentiment analysis: online and social media	Characterize IE	Measure changes in attitudes and beliefs	Low	Low medium	Immediate	Limited, mainly software requirements	Unrepresentative samples, limited to tech-savvy audiences
Online and social media analytics (of DoD messages)	N/A	Measure exposure and reactions (web-based campaigns)	Low	High	Immediate	Limited, mainly software requirements	Only relevant to web-based messages
Informal surveys/ intercept interviews	Test products and characterize IE	Measure attitudes and beliefs	Low	Low	Near term (weeks)	In-house	Not representative, nonresponse and response bias
In-depth interviews	Develop messages	Interpret quantitative results	Medium	Medium	Near term (weeks)	Local researchers or in-house	
Focus groups	Develop messages and test products	Validate and interpret quantitative results	Medium	Medium	Days to months	Local facilitators, often outsourced	Groupthink, difficult to manage, selection bias
Laboratory experiments	Develop messages and theories of change	N/A	Medium high	High	Months	Academic researchers	Requires planning, results can be hard to operationalize
Direct observation and atmospherics	Characterize IE	Measure change in attitudes and beliefs	Medium high	Medium	Days to months	In-house or outsourced	"Signal in noise," no systematic approach
Secondary data/ desk research	Characterize IE and baseline	Measure exposure (e.g., using process similar to Nielsen ratings)	Low	Medium high	Immediate (weeks)	In-house	No control over research design or questions

Measuring Program Processes: Methods and Data Sources

Process evaluation, or program implementation monitoring, seeks to determine the extent to which the program accomplished the tasks it was supposed to accomplish. It is therefore principally concerned with measuring things over which program staff have direct or significant control. Process evaluation is particularly important when a program has failed or fallen short of expectations. If the process evaluation reveals that the program was implemented as planned, it tells the program designers that the theory of change/logic of the effort needs to be revisited, as this would appear to be an instance of potential theory failure rather than program failure.

Process evaluation can be conducted at several points in the campaign process. Specifically, *production* evaluation documents how the message or program was created. *Dissemination* evaluation measures the distribution and placement of messages or the number of events and engagements, depending on the type of campaign.[1] While some researchers include measuring exposure as a component of process evaluation, we address exposure measures separately in this handbook.

The primary sources of data for program implementation measures are direct observation or monitoring of program implementers, media monitoring, service record data, service provider data (e.g., interviews with program managers), and event participant or audience data. When using direct observations, researchers should be sensitive to the "Hawthorne effect" in which subjects are likely to exert extra effort if they are aware they are being observed. Media monitoring should measure message distribution and placement.

Further Reading

In this handbook:

Chapter Five, in the section "Program Failure Versus Theory Failure," discusses potential sources of failure.
Chapter Seven discusses the role of process evaluation in assessment design.

In the accompanying desk reference:

Chapter Nine provides more information on data sources and analysis in the context of DoD IIP assessment efforts.

Measuring Exposure: Measures, Methods, and Data Sources

IIP summative evaluations should include a measure of exposure to campaign materials and several measures that capture the internal processes by which exposure influences behavioral change. Here, we discuss methods for capturing exposure and methods for measuring the internal processes—knowledge, attitudes, and so forth—affected by exposure.

---

[1]   Valente, 2002, pp. 75–77.

The first step in assessing the outcome of an IIP campaign is measuring the extent to which the target audience was *exposed* to the program or message. Program exposure is the degree to which an audience *recalls* and *recognizes* the program:

- *Recall* is measured by unaided or spontaneous questions that ask the respondent in an open-ended manner if he or she had been exposed to the campaign.[2]
- *Format-specific recall* establishes whether the audience member recalls the information from the campaign (e.g., a public service announcement) or from other sources (e.g., state news bulletin).[3]
- *Recognition* is measured by aided or prompted questions that provide a visual or aural cue to assist the respondent in recalling the campaign.[4] Recognition measures have greater response bias.[5]

Recall and recognition measures assess exposure along two dimensions: message awareness—measured by reach, frequency, and recency—and message comprehension:

- *Reach* assesses the number of people who saw or heard the message, and is typically defined as the percentage of the target audience exposed to the message at least once during the campaign.
- *Frequency* measures how often the individuals saw the message, defined as the average number of times a person in the target audience had the opportunity to view the message.[6]
- *Recency* measures are common in IIP evaluation and capture the last time the media was viewed.
- *Comprehension* is the extent to which the audience understood the message.[7]

It is important to avoid making assumptions about exposure based on distribution. For example, a person might be exposed to a radio segment, but that does not mean he or she comprehended the message.[8] What people are actually exposed to is usually a subset of what you put out.[9]

---

[2]   Valente, 2002, p. 184.

[3]   Gerry Power, Samia Khatun, and Klara Debeljak, "'Citizen Access to Information': Capturing the Evidence Across Zambia," in Ingrid Volkmer, ed., *The Handbook of Global Media Research*, Chichester, West Sussex, UK: Wiley-Blackwell, 2012, p. 263.

[4]   Valente, 2002, p. 184.

[5]   Author interview with Ronald Rice, May 9, 2013.

[6]   Author interview with Thomas Valente, June 18, 2013.

[7]   Power, Khatun, and Debeljak, 2012, p. 263.

[8]   Author interview with Gerry Power, April 10, 2013.

[9]   Author interview with Ronald Rice, May 9, 2013.

Content Analysis and Social Media Monitoring

Content analysis involves the systemic observation of traditional press (television, radio, newspaper) and web and social media sources to quantify programs and messages communicated through the media to determine how messages are spreading throughout the target audience. Because media content reflects both dissemination and reactions to the campaign, as well as baseline sentiments, it can be used to inform all three phases of evaluation.

Methods associated with content analysis include *traditional press and broadcast media analysis* (television, radio, newspapers, political events and associated web content) and *social media analysis*. Traditional press and broadcast media analysis is considerably more resource intensive than social media analysis, but, depending on target audience characteristics, it may generate a more representative sample.

Depending on how the information will be used, content analysis must focus on one or both of two issues: (1) the content of interest and (2) the extent to which the sample represents the audience or population of interest. These factors can conflict. For example, social media platforms such as Twitter provide enormous amounts of content that is relatively easy to code, but it is difficult to determine the extent to which the voices generating that content reflect voices within the target audience.

Measuring Observed Changes in Individual and Group Behavior and Contributions to Strategic Objectives

Data on behaviors are difficult to collect in a representative fashion. Nonetheless, to complement and validate self-report measures, the most valid and useful IIP assessments include measures of how the population actually behaves.

Observing Desired Behaviors and Achievement of Influence Objectives

IIP assessment should measure changes in the behavior targeted by the influence objective. For example, if the influence objective is to increase voter turnout, the assessment should measure voter turnout. If the objective is to mislead enemy decisionmaking, the assessment should be capable of capturing the enemy's choices. If the objective is to increase surrenders, surrenders should be tracked over time.

When the behavior cannot be observed systematically or aggregately, researchers can use the participant-observation technique to observe a sample of the target audience. The validity of participant observation is limited by several factors. First, the observer or rater may be biased due to pressures to show program effects. Second, the observer effect biases how the subjects behave when under observation, which is amplified in the case of an armed observer. Third, it is difficult to prove that the sample being observed is representative of the target audience.

Further Reading

In this handbook:

Chapter Seven, in the sections "Designing Valid Assessments" and "Summative Evaluation Design," discusses the difficulty of designing evaluations and isolating the causal role of an IIP effort or campaign from background noise and other variables.

In the accompanying desk reference:

Chapter Nine, in the section "Observing Desired Behaviors and Achievement of Influence Objectives," addresses the use of proxies to measure behaviors that cannot be observed.

Direct and Indirect Response Tracking

In some cases, behaviors can be observed that directly or indirectly gauge the influence of the program because the behaviors can only be reasonably explained by the fact that the audience was exposed to the program. In evaluation research this method is often called *direct response tracking*. For example, a social marketing ad may ask a viewer to undertake a direct and measurable response, such as calling an 800 number or visiting a website. *These are often weak indicators of effects, however, unless research has demonstrated a strong correlation between engaging in the direct response and adopting the desired behavioral change.* To strengthen this approach, some evaluations will use the direct responders for a follow-up evaluation to determine whether and how the information they received shaped their behavior.[10]

Atmospherics and Observable Indicators of Attitudes and Sentiments

If collected and analyzed systematically and rigorously, atmospherics and associated measures can provide more-robust estimates of sentiment than self-report survey data.[11]

---

[10] Coffman, 2002, p. 15.

[11] Author interview with Anthony Pratkanis, March 26, 2013.

Atmospherics is a poorly defined but commonly used term by DoD assessment practitioners. Informally, it refers to a range of observable indicators that are used or could be used to characterize the prevailing mood or atmosphere of the target audience. It is distinguished from large surveys or formal opinion polling research and can gauge sentiments toward U.S. or friendly forces, trust in public institutions, and perceptions of security. Examples include

- how the population responds to patrol vehicles rolling through villages (e.g., throwing stones or cheering)
- the extent to which the population engages with friendly forces (e.g., eye contact, exchanging information, letting friendly forces "in the door")
- the number of people shopping at the bazaar or the traffic on a road used to go to a market
- the number of intelligence tips given to friendly forces by the target audience
- subjective assessment of the mood from trusted local sources through informal interviews.[12]

Because there are a nearly infinite number of possible atmospheric indicators, a central challenge with atmospherics is determining what data are essential to collect and analyze—"finding the signal in the noise." The key, according to the social psychologist and influence expert Anthony Pratkanis, "is coupling those atmospheric measures to objectives."[13] Doing so requires a sophisticated understanding of the cultural context so that evaluators can reliably interpret the meaning behind what they are observing.[14] Researchers should consider using empirical analysis and the Delphi process to determine which atmospheric variables are worth capturing.

While standardization is important, atmospheric measures and data collection strategies also must be flexible enough to be tailored to the local IE and security context. Every locale is potentially different, and indicators will have different meanings depending on the context.

Further Reading

In the accompanying desk reference:

Chapter Nine provides more detail on empirical analysis and the use of the Delphi process to determine which atmospheric variables are worth capturing in the section "Selecting Valid and Useful Atmospheric Measures and Data Sources," along with suggestions systematizing and institutionalizing the collection and analysis of meaningful atmospherics in the section "Improving Atmospheric Data Collection."

---

[12] Author interview on a not-for-attribution basis, December 15, 2013.

[13] Author interview with Anthony Pratkanis, March 26, 2013.

[14] Author interview with LTC Scott Nelson, October 10, 2013.

Aggregate or Campaign-Level Data on Military and Political End States

Another directly observed data source is aggregate data reflecting the extent to which military or political objectives are being achieved. IIP activities should, if the logic model is valid, contribute to the achievement of military and political strategic objectives and end states. For example, if the IIP MOPs suggest that the influence program is working but other indicators suggest that violence is increasing and that the coalition-supported government is losing legitimacy, IIP planners should revisit the logic model, inspect the validity and reliability of their MOPs and MOEs, or both. To track the achievement of broader military and political objectives, IIP assessors should track casualties, recruitment, levels of violence, surrenders, and economic and governance indicators with their area of operations.

Measuring Effects That Are Long-Term or Inherently Difficult to Observe

We have just discussed measures and methods assuming that an outcome has occurred and is observable. However, it is not always the case that the outcome of interest has occurred by the time the assessment must be conducted. A core challenge in IIP assessment is in balancing near-term assessment and reporting requirements with the evaluation efforts and behavioral change over the long term.

Those responsible for evaluating the effectiveness of long-term influence activities commonly find themselves *wishing* that data had been collected historically and over time. To facilitate future longitudinal evaluations, IIP programs need to collect consistent data over time on a broad range of input, output, and outcome variables. Retrospectively collecting or estimating who was engaged and when is expensive and difficult.[15] Because organizations, priorities, and evaluation research questions change over time, it is important to collect data on a wide range of variables that may be relevant to future generations of decisionmakers.[16] Collecting data over long periods is also beneficial because it allows researchers to identify aberrant or unusual waves of data that might suggest cheating or other errors affecting the data collection process.[17]

Further Reading

In this handbook:

Chapter Eight presents more detail on the formative and qualitative research methods covered here.
Chapter Nine discusses the analysis and interpretation of survey data and margins of error, as well as trend analysis and tracking program progress over time.

In the accompanying desk reference:

Chapter Nine, in the section "Analyses and Modeling in Influence Outcome and Impact Evaluation,"

---

[15] Author interview with James Pamment, May 24, 2013.

[16] Author interview with James Pamment, May 24, 2013.

[17] Author interview with Katherine Brown, March 4, 2013.

offers a more detailed discussion of the insights, concepts, and best practices summarized here. The section "Narrative as a Method for Analysis or Aggregation" describes how narratives can support data aggregation.

Chapter Eleven discusses data aggregation in the context of decision support.

Key Takeaways

- *Good data* is not synonymous with *quantitative data*. Depending on the methods and the research question, qualitative data can be more valid, reliable, and useful than quantitative data.
- Exposure should be measured in terms of the audience's ability to recall or recognize a message (e.g., whether they "tuned in"), as opposed to whether they saw it (media impressions).
- Evaluators should not make assumptions about exposure based on distribution. *Reach* is often a misused term in media evaluation.
- Threats to validity associated with self-report measures can be minimized with consistent measurement over time and across areas.
- Good formative research can help determine the relative importance of measuring attitudes versus behaviors because it identifies the extent to which attitudes predict behaviors.
- Content analysis serves many purposes in all three phases of evaluation. In the formative phase, it can characterize the IE and target audience characteristics. In the process phase, it can determine the distribution of the campaign. In the summative phase, it can measure exposure (particularly for web and social media content), as well as reactions and sentiment over time.
- Key challenges with social media analysis are finding the signal in the noise and ensuring that the sample represents the target audience.
- Because there are a nearly infinite number of possible atmospheric indicators, a central challenge with atmospherics is determining what data to collect and analyze.
- Aggregation requires consistent measurement over time and across areas. Consistent, mediocre assessments are better than great, inconsistent assessments.
- The best evaluations triangulate many measures from different methods and data sources. The most valid measures are those that converge across multiple qualitative and quantitative methods.
- The most valid and useful measurements are those that capture trends over time and across areas.

Presenting and Using Assessment

By now, the "spaghetti graph," as it has come to be known, is infamous for its complexity and overlapping lines. According to a *New York Times* article, when General McChrystal was the leader of American and NATO forces in Afghanistan, he jokingly remarked, "When we understand that slide we'll have won the war."[1] The moral of the story is that how one presents and uses assessment matters, because assessment supports decisionmaking, and poorly presented assessments offer poor support to decisionmaking. As Maureen Taylor noted, "The biggest challenge facing assessment is getting information into a form that the people who make decisions on the ground can use."[2]

Assessment and Decisionmaking

As emphasized repeatedly throughout this handbook, assessments should be designed with the needs of stakeholders in mind; this fully carries over to the presentation of assessments. Only by having a clear understanding of both the assessment users (stakeholders, other assessment audiences) and the assessment uses (the purposes served and the specific decisions to be supported) can assessment be tailored in its design and presentation to its intended uses and users and thus adequately support decisionmaking. Presenting information will mean nothing unless the data are shared with stakeholders who play a major role in decisionmaking. This provides an impetus to offer better training in data-driven decisionmaking and to make the results and data more accessible to those not trained in research methods.[3]

The Presentational Art of Assessment Data

Deciding how and how much assessment data to present in a report or briefing is a difficult challenge. Too much data, and the reader or recipient will drown

---

[1] Elisabeth Bumiller, "We Have Met the Enemy and He Is PowerPoint," *New York Times*, April 26, 2010.

[2] Author interview with Maureen Taylor, April 4, 2013.

[3] Author interview with Maureen Taylor, April 4, 2013.

in the data, fail to see the forest for the trees, or simply ignore the material as being too opaque and not sufficiently accessible. Too little data, on the other hand, and the recipient will lack confidence in the results, question the validity of findings, or ask important questions that the underlying (but unavailable) data should easily answer.

When presenting data in charts and graphs, consider the most effective way to appropriately communicate the information to the audience. Before constructing charts and graphs, consider their necessity and structure. Reduce "chart junk," including unnecessary graphics. Be thoughtful when ordering data points; for example, figure out whether to rank points in order of priority or whether alphabetical order is appropriate.[4] Overall, it is best to present dense and rich data as clearly and simply as possible to let the research speak for itself. However, do not assume that data speak for themselves; what is obvious to an assessor who has spent hours poring over and analyzing a matrix of data will likely not be obvious to a first-time viewer of even a relatively simple data table.

As the example of General McChrystal's spaghetti graph demonstrates, Power-Point has its own limitations. In an article titled "PowerPoint Is Evil," Edward Tufte, a famed researcher on the visual presentation of data, wrote, "The practical conclusions are clear. PowerPoint is a competent slide manager and projector. But rather than supplementing a presentation, it has become a substitute for it. Such misuse ignores the most important rule of speaking: Respect your audience."[5] While many IIP assessment presentations and briefings must still rely on PowerPoint, the takeaway remains clear: Understand and meet the needs of your audience, and respect your audience. Make it clear when complicated data support a simple conclusion, and have a more detailed presentation of those data available if needed (perhaps in the backup slides). Again, Tufte's words are instructive:

> Presentations largely stand or fall on the quality, relevance, and integrity of the content. If your numbers are boring, then you've got the wrong numbers. If your words or images are not on point, making them dance in color won't make them relevant. Audience boredom is usually a content failure, not a decoration failure.[6]

One form that can be very effective is quantitative data supported by narrative and qualitative data. Qualitative data are illustrative and provide context to the numbers, while narrative is a strong way to summarize assessments. To be sure, those narratives that explicitly mention a theory of change/logic of the effort and how well it is working are even better. All assessments—even narratives—should clarify the underlying data and level of confidence in the result. Presentational art includes finding

---

[4]   Howard Wainer, "How to Display Data Badly," *American Statistician*, Vol. 38, No. 2, May 1984.

[5]   Edward Tufte, "PowerPoint Is Evil," *Wired*, September 2003.

[6]   Tufte, 2003.

the right balance in discussing methods and evidence. As one SME concluded, "It is important that you do good science; it is also important that you sell good science."[7]

Tailor Presentation to Stakeholders

When presenting data, knowing your audience is paramount. Henry May articulates three principles for tailoring data presentation to specific audiences:

- *Understandability:* Results need to be reported in a form that can be widely understood, makes minimal assumptions about the audience's familiarity with statistics, and avoids the overuse of jargon.
- *Interpretability:* The metric or unit of measure must be easily explained.
- *Comparability:* Statistics can be compared directly, obviating any need for further manipulation.[8]

Commanders and decisionmakers are inundated with more data than they can reasonably comprehend, so the onus is on those *presenting* the data to tailor their presentations to stakeholders. We've all heard of the perfect "elevator speech," or the 30-second pitch that perfectly captures the main takeaways from your research. Tailoring presentations to stakeholders is built around this same logic.

Dissemination should adhere to a certain framework, and findings need to be tailored to their intended audiences.[9] Decisionmakers in conflict zones are busy. In terms of reading evaluations, the executive summary is critical: "Often, no one reads anything except the executive summary, so you have to make it count."[10]

Finally, to properly tailor the presentation of assessment results to stakeholders, it is crucial to know what they need to know to support the decisions they need to make. Here, it is important to take care when aggregating assessments of individual efforts or programs. In other words, sometimes the whole really is greater than the sum of its parts.

How to Present Data, and How Much

Closely related to tailoring presentations to stakeholders is the question of how much data to present and in what format. Part of any effective assessment will include communicating progress (or a lack thereof) in both interim and long-term measures. Some stakeholders will need more hand-holding than others, but the onus is on the research organization to have the data and the ability to provide updates in a meaningful and

---

[7]   Author interview on a not-for-attribution basis, July 30, 2013.

[8]   Henry May, "Making Statistics More Meaningful for Policy Research and Program Evaluation," *American Journal of Evaluation*, Vol. 25, No. 4, 2004.

[9]   Author interview with Thomas Valente, June 18, 2013.

[10]   Author interview with Amelia Arsenault, February 14, 2013.

measurable way.[11] NATO's Joint Analysis and Lessons Learned Centre framework for the evaluation of public diplomacy has devised three separate evaluation products to represent three levels of reporting: dashboards, scorecards, and evaluation reports.

A dashboard provides an overview of monitoring, usually of outputs. It can be used in real time with some media-monitoring applications and can be used to produce regular and frequent reports. A dashboard is essentially data with little or no built-in evaluation and limited explanative narrative. A dashboard would typically be updated at least monthly.

A scorecard is a display format for less frequent reporting, as it shows progress toward the desired outcomes and desired impacts. A scorecard is essentially data with little or no bulletin evaluation and limited explanative narrative. A scorecard would typically be updated quarterly or biannually.

An evaluation report is a periodic, typically annual, evaluation of results. It presents a balanced view of all relevant results and aims to show what meaningful changes have occurred and how they might be linked to activities, and the evaluation judges whether the objectives have been achieved. It should contain narrative answers to the research questions and explain what has worked, what hasn't, and, whenever possible, why. Evaluation reports can also be published to cover a specific event or program.[12]

Data Visualization

Assessments can be presented in a variety of forms, including research reports, policy memorandums, and PowerPoint briefings packed with a dizzying array of quantitative graphs, maps, and charts. Senior military leaders and policy staffs use these materials for a variety of purposes, including to assess the progress of military campaigns, allocate (or reallocate) resources, identify trends that may indicate success or failure, and discern whether and when it may be necessary to alter a given strategy.[13] As such, it is important to think about different ways to present important data so that they can be visualized properly and have the proper effect.

Sometimes, to truly make sense of the data, it is important to visualize them. To really ramp up the productivity of the data, you need a way to ramp up the visualization technology. There are a number of software solutions that can support more-complicated or multidimensional displays of data; one such software program is called Ignite. This program, and others like it, allows you to visualize structured and unstructured data. If data lend themselves to more-complex visual presentations, then using this type of program can be a great way to demonstrate progress toward your end

---

[11] Author interview with Heidi D'Agostino and Jennifer Gusikoff, March 2013.

[12] NATO, Joint Analysis and Lessons Learned Centre, 2013, p. 12. Illustrations of each type of evaluation product are provided in chapter 3 of the framework.

[13] Connable, 2012, p. iii.

state.[14] These infographics can also help communicate research results to decisionmakers in the field.[15] A picture is indeed worth a thousand words, if you can generate the right picture.

The Importance of Narratives

While visual representations of data can help communicate key points to an audience, to avoid losing the nuance of assessment results, it is important to place metrics in context and frame these visual representations within broader explanatory narratives. This means balancing quantitative metrics with probability and accuracy ratings and also identifying and explaining gaps in the available information. To remain transparent, all information should be clearly sourced. Quantitative reports should be presented as part of holistic, all-source analysis as part of a narrative.[16]

Narratives are even more effective if they make explicit reference to a theory of change/logic of the effort, explain critical nodes and assumptions, and combine quantitative data with anecdotes to color and provide context to the numbers.[17] Depending on the audience, the use of strong anecdotes, such as messages illustrating adversary awareness of and concern about an IIP effort, can be a potent demonstration of the effectiveness of a campaign. The following sections address the benefits of narratives in increasing understanding, which facilitates the translation of aggregated data into terms that best support decisionmaking and the process of soliciting valuable feedback from end users of assessment results.

Aggregated Data

Transparency and analytic quality might enhance the credibility of aggregated quantitative data.[18] It is important to remember that ordinal scales (scales with entries reporting order or ranking, but not necessarily uniform distance between ordered or ranked items) can be aggregated and summarized with narrative expressions but not (accurately) with numbers. The simple statement "All subordinate categories scored a B or above except for reach in the Atlantica region, which scored a D," is much more informative than "The Atlantica region scored a 2.1 for reach."

Because a whole really can be greater than the sum of its parts, one must take great care when aggregating assessments of individual efforts or programs to avoid junk arithmetic. Ordinal scales are better represented as letter grades than as numbers; it is

---

[14] Author interview with LTC Scott Nelson, October 10, 2013.

[15] Author interview with Gerry Power, April 10, 2013.

[16] Connable, 2012, p. xix.

[17] Author interview with Maureen Taylor, April 4, 2013.

[18] Connable, 2012, p. xix.

harder to inappropriately average C, C, and A than it is to inappropriately average 1, 1, and 4. Ordinal scales can be aggregated and summarized with narratives but not with numerical averages.

Report Assessments and Feedback Loops

Disseminating findings is just one piece of the puzzle when it comes to supporting decisionmaking. To generate valuable feedback loops, those preparing and contributing to the research must receive feedback from the end user (the stakeholder or decisionmaker). Efforts to improve transparency should include stressing the importance of feedback, both from individuals who have a broad understanding of the issue of interest and from those who have an understanding of specific circumstances and audiences.

Evaluating Evaluations: Meta-Analysis

With all of the time, effort, and resources dedicated to conducting evaluations, how do we know whether an evaluation is sound? By stepping back and conducting research about research, we are, in essence, conducting a form of meta-analysis. In the evaluation context, this means using metaevaluation to *assess the assessment*. Metaevaluation is the extent to which the quality of the evaluation itself is assured and controlled. Its purpose is to be responsive to the needs of its intended users and to identify and apply appropriate standards of quality. Metaevaluations should be based on adequate and accurate documentation.

Further Reading

The metaevaluation checklist that accompanies this handbook online is designed for assessments of actual influence efforts (though not for supporting or enabling efforts that do not have some form of influence as an outcome).

In the accompanying desk reference:

Chapter Nine, in the section "Narrative as a Method for Analysis or Aggregation," discusses the role of narrative in facilitating aggregation.

Chapter Eleven offers more detail on metaevaluation approaches, as well as quality indexes for evaluation design.

Key Takeaways

- Tailor the presentation of assessment results to the stakeholder. Those preparing assessments should be asking, "What do stakeholders need to know to support the decisions they need to make?" Not every stakeholder wants or needs a report, and not every stakeholder wants or needs a briefing.
- Quantitative data supported by qualitative data can be very effective: The combination can help illustrate findings and provide context for the numbers.
- Narratives can be an excellent way to summarize assessment results, and those that explain the attendant theory of change and how well it is working in a nuanced

context are even better. All assessments should make clear what data form their foundation and how confident stakeholders should be of the results.

- Building on the previous point, narratives also support data aggregation and the process of soliciting feedback from end users of assessment results by increasing stakeholder and decisionmaker understanding of what might be complex or opaque approaches to rolling up quantitative data.
- Stakeholders are not the only ones who stand to benefit from assessment data. Input, feedback, and guidance derived from the results should be shared with those who have contributed to the assessment process, as well as, when possible, those who are working on similar efforts.
- Assessors need to take care when aggregating assessments of individual efforts or programs. Sometimes, the whole really is greater than the sum of its parts. The metaevaluation checklist that accompanies this handbook online can be an effective tool for assessing assessments.

Developing a Culture of Assessment

Organizations that do assessment well usually have a culture that values assessment. Without an understanding and appreciation for what assessment can accomplish, it is much easier to dismiss assessment as an afterthought. A critical component to conducting assessment—albeit a component that is often underappreciated—is building organizations that value *research*.

This topic is covered at great length in the accompanying desk reference, but it is so central for shaping the high-level decisionmaking that DoD IIP assessment supports that we have elected to emphasize it here as well. For the background of practitioners who are part of the larger DoD organizational structure and whose contributions to larger campaigns and to DoD initiatives writ large, we present here the broad characteristics of organizations with effective assessment cultures.

- Organizations that do assessment well usually have organizational cultures that value assessment.
- Assessment requires resources (as a rule of thumb, roughly 5 percent of program resources should be dedicated to assessment).
- Successful assessment depends on the willingness of leadership to learn from the results. (This echoes the admonition in Chapter Two's discussion of operational design in JP 5-0 for leaders to promote and embrace constant change, learning, and adaptation.)
- Assessment requires data to populate measures—and intelligence is potentially a good data source.
- IIP efforts should be broadly integrated into DoD processes, and IIP assessment should be integrated with broader DoD assessment efforts. There remains a gap in doctrinal focus on assessment; this is why we point out throughout this handbook where observed strong practices would conform to JOPP guidance.
- Assessment needs advocacy, improved doctrine and training, more trained personnel, and greater access to assessment and influence expertise to break the current "failure cycle" for assessment in DoD.
- Independent assessment and formal devil's advocacy are valuable tools in promoting a culture of assessment, especially in avoiding rose-tinted glasses

in understanding the operational environment. These approaches could be incorporated into JOPP during COA analysis and war-gaming, but they should also be included in the iterative cycle of operational design.

- Assessment starts in planning and continues through execution. Overlaying the JOPP steps, this means assessment begins with mission analysis (step 2) and continues through to step 7—plan or order development.

When organizing for assessment, IIP should be broadly integrated into DoD routine processes as well as within broader DoD assessment. With IIP assessment, there is often a lack of shared understanding about the logic of effort and the assessment process, so there is a need to be much more explicit about all the steps and assumptions. Some best practices more generally include making sure that assessors are independent enough (and brave enough) to identify and decry problems in execution or assumptions when evaluation reveals them, to avoid overoptimism through independence or formal devil's advocacy, and to not be afraid to collaborate with experts from social science or behavioral communication.

One of our key pieces of advice to DoD leaders is this: Don't fear bad news. No organization—not even the most transparent—refrains from cringing just a little bit when its daily activities are placed under a microscope. However, an organization that has developed an assessment culture will be more accepting of bad news and will welcome it as an opportunity to improve and learn.

Further Reading

In this handbook:

Chapter Three, in Box 3.2, "Challenge: Lack of Shared Understanding," highlights the importance and challenges of building a shared understanding of IRCs. The chapter also touches on this issue in the section "Requirement 1: Congressional Interest and Accountability," as it relates to congressional stakeholders.

In the accompanying desk reference:

Chapter Four covers the full range of topics associated with organizing for assessment and the challenges involved in doing so.

Conclusions and Recommendations

This handbook was designed to be an easy-to-navigate, quick-reference guide to planning and conducting assessments of DoD IIP efforts, analyzing the data generated, and presenting the results to decisionmakers and stakeholders. It also offers some background on current assessment practices in DoD and the typical users and uses of DoD IIP assessment results. Each chapter has its own summary that lists the key insights and takeaways from the discussion it contains. These final conclusions reprise only the most essential of these numerous insights, those that are most intimately connected with the report's recommendations.

Key Conclusions

- If the prospects for an effort are uncertain, *fail fast* by rapidly trying, assessing, and adjusting the effort until it either works or needs to be abandoned.
- Formative, process, and summative evaluations have nested and connected relationships; unexpected poor performance at higher levels can be explained by thoughtful assessment at lower levels. This is captured in the hierarchy of evaluation. (See **Chapter Three**, "Three Types of Evaluation: Formative, Process, and Summative.")
- Good assessment supports and informs decisionmaking. Assessments need to be tailored to the needs of end users in both their design and their presentation. (See **Chapter Three**, "Uses and Users of Assessment.")
- Good objectives are "SMART": specific, measurable, achievable, relevant, and time-bound. (See **Chapter Four**, "Characteristics of SMART or High-Quality Objectives.")
- When the program does not produce all the expected outcomes and one wants to determine why, a logic model (or other articulation of the theory of change/logic of the effort) really shines. (See **Chapter Five**, "Building a Logic Model or Theory of Change.")
- Good measures are valid, reliable, feasible, and useful. (See **Chapter Six**, "Attributes of Good Measures.")
- To balance the strengths and weaknesses across different designs, the best evaluations draw from a compendium of studies with multiple designs and methods that converge on key results. (See **Chapter Seven**, "The Best
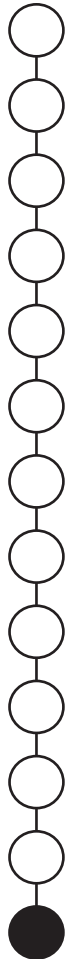
Evaluations Draw from a Compendium of Studies with Multiple Designs and Approaches.")

- The plural of *anecdote* is not data. Qualitative data should be generated by rigorous social science methods. Likewise, decisionmakers should not be expected to make decisions on the basis of a single quantitative method. (See **Chapter Eight**, "The Importance and Role of Qualitative Research Methods.")
- Nonresponse and lack of access are challenges inherent in all survey efforts. This is especially true for survey efforts conducted in conflict environments, where populations may move frequently, people may lack access to telephones or the Internet, and areas are inaccessible. (See **Chapter Nine**, "Challenges to Survey Sampling.")
- The best evaluations triangulate many measures from different methods and data sources. The most valid measures are those that converge across multiple qualitative and quantitative methods. (See **Chapter Ten**, "Overview of Research Methods for Evaluating Influence Effects.")
- Narratives can be an excellent way to summarize and aggregate assessment results, and those that include the attendant theory of change/logic of the effort and how well it is working in a nuanced context are even better. (See **Chapter Eleven**, "The Importance of Narratives.")
- Organizations that do assessment well usually have cultures that value assessment. (See **Chapter Twelve**.)

Recommendations

This handbook contains insights that are particularly useful for those charged with planning and conducting assessment; the companion volume, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Desk Reference*, offers an abundance of information that is relevant to other stakeholders, including those who make decisions based on assessments and those responsible for setting priorities and allocating resources for assessment and evaluation.[1]

Our recommendations for assessment practitioners echo some of the most important practical insights described in the key takeaways at the end of each chapter and the summary conclusions at the end of this handbook:

- *Demand SMART objectives.* Where program and activity managers cannot provide assessable objectives, assessment practitioners should infer or create their own.
- *Be explicit about theories of change.* The theory of change or logic of the effort ideally comes from the commander or program designers, but, if the logic of the

---

[1]   Paul et al., 2015a.

effort is not made explicit, assessment practitioners should elicit or develop one in support of assessment.

- *Insist that resources are provided for assessment.* Assessment is not free, and if its benefits are to be realized, it must be resourced. Presenting assessment results in ways that are tailored to specific stakeholders, highlighting successes in saving time and resources, and ensuring that data collection, measures, and results are as transparent as possible will help gain buy-in from stakeholders and DoD leadership.
- *Take care to match the design, rigor, and presentation of assessment results to the intended uses and users.* Assessment supports decisionmaking, and providing the best decision support possible should remain at the forefront of practitioners' minds. The ways in which assessment results will be used by decisionmakers must be a consideration throughout the assessment process. This may involve some amount of prediction, as decisionmakers may not always know what information they require, and it can be time-consuming and expensive to assemble the results required after data have been collected.

Practitioners depend to a great extent on leadership support and shared understanding with stakeholders and decisionmakers, just as leadership and stakeholders depend on practitioner understanding of their needs and resource constraints. As such, we reiterate here some recommendations for the broader DoD IIP community, including stakeholders, proponents, and capability managers for IO, public affairs, military information support operations, and all other information-related capabilities. The following recommendations, drawn primarily from points in *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Desk Reference* but also addressed to some extent in this handbook, emphasize how advocacy and a few specific practices can improve the quality and use of assessment results across the community:

- DoD leadership needs to *provide greater advocacy, better doctrine and training, and improved access to expertise (in both influence and assessment)* for DoD IIP assessment efforts. Assessment is important for both accountability and improvement, and it needs to be treated as such.
- DoD doctrine needs to *establish common assessment standards.* There is a large range of possible approaches to assessment, with a similarly large range of possible assessment rigor and quality. The routine and standardized employment of something like the assessment metaevaluation checklist that accompanies this handbook online would help ensure that all assessments meet a target minimum threshold.
- DoD leadership and guidance need to recognize that *not every assessment must be conducted to the highest standard.* Sometimes, good enough really is good enough, and significant assessment expenditures cannot be justified for some efforts, either

because of the low overall cost of the effort or because of its relatively modest goals.

- DoD should *conduct more formative research.* Formative research can improve IIP efforts and programs and facilitate the assessment process. We offer the following specific recommendations:
  - Conduct target audience analysis with greater frequency and intensity, and improve capabilities in this area.
  - Conduct more pilot testing, more small-scale experiments, and more early efforts to validate a specific theory of change in a new cultural context.
  - Try different things on small scales to learn from them (i.e., fail fast).
  - DoD leaders need to *explicitly incorporate assessment into orders.* If assessment is in the operation order, the execute order, or even a fragmentary order, then it is clearly a requirement and will be more likely to occur, with requests for resources or assistance less likely to be resisted.
  - DoD leaders should support the *development of a clearinghouse of validated (and rejected) IIP measures.* When it comes to assessment, the devil is in the details. Even when assessment principles are adhered to, some measures just do not work out, either because they prove hard to collect or because they end up being poor proxies for the construct of interest. Assessment practitioners should not have to develop measures in a vacuum. A clearinghouse of measures tried (with both success and failure) would be an extremely useful resource.

References

Altai Consulting, "Afghan Media in 2010," prepared for the U.S. Agency for International Development, 2010. As of October 22, 2014:
http://www.altaiconsulting.com/library_details/160986

Arsenault, Amelia, Sheldon Himelfarb, and Susan Abbott, *Evaluating Media Interventions in Conflict Countries*, Washington, D.C.: United States Institute of Peace, 2011. As of August 22, 2014:
http://www.usip.org/sites/default/files/resources/PW77.pdf

Babbie, Earl, *Survey Research Methods*, 2nd ed., Belmont, Calif.: Wadsworth Publishing Company, 1990.

Becker, David C., and Robert Grossman-Vermaas, "Metrics for the Haiti Stabilization Initiative," *Prism*, Vol. 2, No. 2, March 2011, pp. 145–158.

Brick, J. M., and G. Kalton, "Handling Missing Data in Survey Research," *Statistical Methods in Medical Research*, Vol. 5, No. 3, September 1996, pp. 215–238.

British Council, *Trust Pays: How International Cultural Relationships Build Trust in the UK and Underpin the Success of the UK Economy*, Edinburgh, UK, 2012. As of October 17, 2014:
http://www.britishcouncil.org/sites/britishcouncil.uk2/files/trust-pays-report.pdf

Bruce, Sarah, and Mary Tiger, *A Review of Research Relevant to Evaluating Social Marketing Mass Media Campaigns*, Durham, N.C.: Clean Water Education Partnership, undated.

Bulmer, Martin, "Introduction: The Problem of Exporting Social Survey Research," *American Behavioral Scientist*, Vol. 42, No. 2, October 1998, pp. 153–167.

Bumiller, Elisabeth, "We Have Met the Enemy and He Is PowerPoint," *New York Times*, April 26, 2010.

Coffman, Julia, *Public Communication Campaign Evaluation*, Washington, D.C.: Communications Consortium Media Center, May 2002.

Connable, Ben, *Embracing the Fog of War: Assessment and Metrics in Counterinsurgency*, Santa Monica, Calif.: RAND Corporation, MG-1086-DOD, 2012. As of August 22, 2014:
http://www.rand.org/pubs/monographs/MG1086.html

Crano, William D., and Marilynn B. Brewer, *Principles and Methods of Social Research*, 2nd ed., Mahwah, N.J.: Lawrence Erlbaum Associates, 2002.

Department of General Psychology and Occupational Psychology, Heinrich-Heine-Universität Düsseldorf, "G*Power: Statistical Power Analyses for Windows and Mac," web page, undated. As of August 22, 2014:
http://www.gpower.hhu.de

DeVellis, Robert F., *Scale Development: Theory and Applications*, 3rd ed., Thousand Oaks, Calif.: Sage Publications, 2012.

Dillman, Don A., Jolene D. Smyth, and Leah Melani Christian, *Internet, Mail, and Mixed-Mode Surveys: The Tailored Design Method*, 3rd ed., Hoboken, N.J.: John Wiley and Sons, 2009.

Downes-Martin, Stephen, "Operations Assessment in Afghanistan Is Broken: What Is to Be Done?" *Naval War College Review*, Vol. 64, No. 4, Fall 2011, pp. 103–125.

Eles, P. T., E. Vincent, B. Vasiliev, and K. M. Banko, *Opinion Polling in Support of the Canadian Mission in Kandahar: A Final Report for the Kandahar Province Opinion Polling Program, Including Program Overview, Lessons, and Recommendations*, Ottawa, Ont.: Defence R&D Canada, Centre for Operational Research and Analysis, DRDC CORA TR 2012-160U, September 2012.

Faul, Franz, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang, "Statistical Power Analyses Using G*Power 3.1: Tests for Correlation and Regression Analyses," *Behavior Research Methods*, Vol. 41, No. 4, November 2009, pp. 1149–1160.

Fowler, Floyd J., *Improving Survey Questions: Design and Evaluation*, Thousand Oaks, Calif.: Sage Publications, 1995.

Grier, Sonya, and Carol A. Bryant, "Social Marketing in Public Health," *Annual Review of Public Health*, Vol. 26, 2005, pp. 319–339.

Haims, Marla C., Melinda Moore, Harold D. Green, and Cynthia Clapp-Wincek, *Developing a Prototype Handbook for Monitoring and Evaluating Department of Defense Humanitarian Assistance Projects*, Santa Monica, Calif.: RAND Corporation, TR-784-OSD, 2011. As of August 22, 2014: http://www.rand.org/pubs/technical_reports/TR784.html

Hayden, Craig, "Another Perspective on IIP Social Media Strategy," *Intermap*, July 23, 2013. As of August 22, 2014: http://intermap.org/2013/07/23/another-perspective-on-iip-social-media-strategy

Headquarters, U.S. Department of the Army, *Inform and Influence Activities*, Field Manual 3-13, Washington, D.C., January 2013a.

———, *Military Information Support Operations*, Field Manual 3-53, Washington, D.C., January 2013b.

Headquarters, U.S. Department of the Army, and Headquarters, U.S. Marine Corps, *Psychological Operations, Tactics, Techniques, and Procedures*, Field Manual 3-05.301/Marine Corps Reference Publication 3-40.6A, Washington, D.C., December 2003.

Hubbard, Douglas W., *How to Measure Anything: Finding the Value of "Intangibles" in Business*, Hoboken, N.J.: John Wiley and Sons, 2010.

The Initiatives Group, *Information Environment Assessment Handbook*, version 2.0, Washington, D.C.: Office of the Under Secretary of Defense for Intelligence, 2013.

Jeffery, Mark, *Data-Driven Marketing: The 15 Metrics Everyone in Marketing Should Know*, Hoboken, N.J.: John Wiley and Sons, 2010.

Joint Information Operations Warfare Center, *Joint Information Operations Assessment Framework*, October 1, 2012.

Ketchum Global Research and Analytics, *The Principles of PR Measurement*, undated. As of August 22, 2014: http://www.ketchum.com/sites/default/files/principles_of_pr_measurement_0.pdf

LaRivee, Dave, *Best Practices Guide for Conducting Assessments in Counterinsurgencies*, Washington, D.C.: U.S. Air Force Academy, December 2011.

May, Henry, "Making Statistics More Meaningful for Policy Research and Program Evaluation," *American Journal of Evaluation*, Vol. 25, No. 4, 2004, pp. 525–540.

McGuire, William J., "McGuire's Classic Input-Output Framework for Constructing Persuasive Messages," in Ronald Rice and Charles Atkin, eds., *Public Communication Campaigns*, Thousand Oaks, Calif.: Sage Publications, 2012, pp. 143–146.

Mertens, Donna M., and Amy T. Wilson, *Program Evaluation Theory and Practice: A Comprehensive Guide*, New York: Guilford Press, 2012.

Muñoz, Arturo, *U.S. Military Information Operations in Afghanistan: Effectiveness of Psychological Operations 2001–2010*, Santa Monica, Calif.: RAND Corporation, MG-1060-MCIA, 2012. As of August 22, 2014:
http://www.rand.org/pubs/monographs/MG1060.html

NATO—*See* North Atlantic Treaty Organization.

North Atlantic Treaty Organization, Joint Analysis and Lessons Learned Centre, *A Framework for the Strategic Planning and Evaluation of Public Diplomacy*, Lisbon, Portugal, June 2013.

Office of the Inspector General, U.S. Department of State, *Inspection of the Bureau of International Information Programs*, May 2013. As of August 22, 2014:
http://oig.state.gov/documents/organization/211193.pdf

Paul, Christopher, *Strategic Communication: Origins, Concepts, and Current Debates*, Santa Barbara, Calif.: Praeger, 2011.

Paul, Christopher, Harry J. Thie, Elaine Reardon, Deanna Weber Prine, and Laurence Smallman, *Implementing and Evaluating an Innovative Approach to Simulation Training Acquisitions*, Santa Monica, Calif.: RAND Corporation, MG-442-OSD, 2006. As of August 22, 2014:
http://www.rand.org/pubs/monographs/MG442.html

Paul, Christopher, Jessica Yeats, Colin P. Clarke, and Miriam Matthews, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Desk Reference*, Santa Monica, Calif.: RAND Corporation, RR-809/1-OSD, 2015a. As of April 2015:
http://www.rand.org/pubs/research_reports/RR809z1.html

———, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: An Annotated Reading List*, Santa Monica, Calif.: RAND Corporation, RR-809/3-OSD, 2015b. As of April 2015:
http://www.rand.org/pubs/research_reports/RR809z3.html

Power, Gerry, Samia Khatun, and Klara Debeljak, "'Citizen Access to Information': Capturing the Evidence Across Zambia," in Ingrid Volkmer, ed., *The Handbook of Global Media Research*, Chichester, West Sussex, UK: Wiley-Blackwell, 2012. As of August 22, 2014:
http://www.intermedia.org/wp-content/uploads/Chapter-15-Global-Media-Research1.pdf

Rice, Ronald E., and Dennis R. Foote, "A Systems-Based Evaluation Planning Model for Health Communication Campaigns in Developing Countries," in Ronald E. Rice and Charles K. Atkin, eds., *Public Communication Campaigns*, 4th ed., Thousand Oaks, Calif.: Sage Publications, 2013, pp. 69–82.

Rosenthal, Robert, and Ralph L. Rosnow, *Essentials of Behavioral Research*, 3rd ed., New York: McGraw-Hill, 2008.

Rossi, Peter H., Mark W. Lipsey, and Howard E. Freeman, *Evaluation: A Systematic Approach*, Thousand Oaks, Calif.: Sage Publications, 2004.

Schaefer, Joseph L., and John W. Graham, "Missing Data: Our View of the State of the Art," *Psychological Methods*, Vol. 7, No. 2, June 2002, pp. 147–177.

Schroden, Jonathan, "Why Operations Assessments Fail: It's Not Just the Metrics," *Naval War College Review*, Vol. 64, No. 4, Fall 2011, pp. 89–102.

Schwarz, Norbert, and Gerald L. Clore, "Mood, Misattribution, and Judgments of Well-Being: Information and Directive Functions of Affective States," *Journal of Personality and Social Psychology*, Vol. 45, No. 3, September 1983, pp. 513–523.

SenseMaker, homepage, undated. As of October 20, 2014:
http://www.sensemaker-suite.com

Stanaland, Andrea J. S., and Linda L. Golden, "Consumer Receptivity to Social Marketing Information: The Role of Self-Rated Knowledge and Knowledge Accuracy," *Academy of Marketing Studies Journal*, Vol. 13, No. 2, 2009, pp. 26–32.

Taylor, Maureen, "Methods of Evaluating Media Interventions in Conflict Countries," paper prepared for the workshop "Evaluating Media's Impact in Conflict Countries," Caux, Switzerland, December 13–17, 2010. As of August 22, 2014:
http://www.global.asc.upenn.edu/fileLibrary/PDFs/taylorcaux2.pdf

Thornton, Ronald J., "Likert Scales: An Assessment Application," *IO Sphere*, Summer 2013.

Tufte, Edward "PowerPoint Is Evil," *Wired*, September 2003. As of August 22, 2014:
http://www.wired.com/wired/archive/11.09/ppt2.html

UK Polling Report, "Support for the Iraq War," online database, undated. As of August 22, 2014:
http://ukpollingreport.co.uk/iraq

Upshur, William P., Jonathan W. Roginski, and David J. Kilcullen, "Recognizing Systems in Afghanistan: Lessons Learned and New Approaches to Operational Assessments," *Prism*, Vol. 3, No. 3, 2012, pp. 87–104.

U.S. Agency for International Development, "Logical Framework Template: Basic," web page, undated. As of October 20, 2014:
http://usaidprojectstarter.org/content/logical-framework-template-basic

U.S. Joint Chiefs of Staff, *Joint Operation Planning*, Joint Publication 5-0, Washington, D.C., August 11, 2011a.

———, *Commander's Handbook for Assessment Planning and Execution*, version 1.0, Suffolk, Va.: Joint and Coalition Warfighting, September 9, 2011b.

U.S. Marine Corps, *Assessment: MAGTF Staff Training Program (MSTP)*, MSTP Pamphlet 6-9, Quantico, Va.: Marine Air-Ground Task Force Staff Training Program, October 25, 2007.

Valente, Thomas W., *Evaluating Health Promotion Programs*, Oxford, UK: Oxford University Press, 2002.

Wainer, Howard, "How to Display Data Badly," *American Statistician*, Vol. 38, No. 2, May 1984, pp. 137–147.

Yeats, Jessica M., and Walter L. Perry, *Review of the Regional Center Enterprise Measures of Effectiveness Plan*, unpublished RAND research, 2011.

To achieve key national security objectives, the U.S. government and the U.S. Department of Defense (DoD) must communicate effectively and credibly with a broad range of foreign audiences. DoD spends more than $250 million per year on inform, influence, and persuade (IIP) efforts, but how effective (and cost-effective) are they? How well do they support military objectives? Could some of them be improved? If so, how? DoD has struggled with assessing the progress and effectiveness of its IIP efforts and in presenting the results of these assessments to stakeholders and decisionmakers. To address these challenges, a RAND study compiled examples of strong assessment practices across sectors, including defense, marketing, public relations, and academia, distilling and synthesizing insights and advice for the assessment of DoD IIP efforts and programs. This handbook was designed to be an easy-to-navigate, quick-reference guide to planning and conducting assessments of DoD IIP efforts, analyzing the data generated, and presenting the results. It also offers some background on current assessment practices in DoD and the typical users and uses of DoD IIP assessment results. A companion volume, *Assessing and Evaluating Department of Defense Efforts to Inform, Influence, and Persuade: Desk Reference*, offers a more detailed exploration and additional examples of assessment in practice.

RAND | NATIONAL DEFENSE RESEARCH INSTITUTE

www.rand.org

$28.50

52850

9 780833 088970